

DIE EXPERIMENTELLE EINZELFALLUNTERSUCHUNG
IN DER ERZIEHUNGSWISSENSCHAFTLICHEN FORSCHUNG

I n a u g u r a l - D i s s e r t a t i o n

Zur Erlangung des Doktorgrades

der philosophischen Fakultät

der Universität Düsseldorf

vorgelegt von

Hsen-hsing Ma

aus Tainan, Taiwan, China

Düsseldorf 1979



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41544304_0035

DIE EXPERIMENTELLE EINZELFALLUNTERSUCHUNG
IN DER ERZIEHUNGSWISSENSCHAFTLICHEN FORSCHUNG

I n a u g u r a l - D i s s e r t a t i o n

Zur Erlangung des Doktorgrades

der philosophischen Fakultät

der Universität Düsseldorf

vorgelegt von

Hsen-hsing Ma

aus Tainan, Taiwan, China

Düsseldorf 1979

D 61



Gedruckt mit Genehmigung der philosophischen Fakultät
der Universität Düsseldorf

Referent: Prof. Dr. V. Krumm

Korreferent: Prof. Dr. H. Nickel

Tage der mündlichen Profung:

10. 12. 1979	Erziehungswissenschaft	Prof. Dr. V. Krumm
11. 12. 1979	Psychologie	Prof. Dr. H. Nickel
20. 12. 1979	Soziologie	Prof. Dr. R. Münch

Die experimentelle Einzelfalluntersuchung
in der erziehungswissenschaftlichen Forschung

Inhaltsverzeichnis

	Seite
1. Einleitung	1
1.1 Theorie-Praxis-Verhältnis	2
1.2 Welche Forschungsstrategie ist für praxisorientierte Experimente angemessen?	8
2. Untersuchung I: Die gegenwärtige erziehungswissenschaftliche Forschungslage in Deutschland	14
2.1 Fragestellung	14
2.2 Methode	14
2.2.1 Auswahl der Zeitschriften	14
2.2.2 Auswahl der Artikel	16
2.2.3 Sammlung der Daten	17
2.3 Ergebnisse und Diskussion	17
3. Die Validität der experimentellen Einzelfalluntersuchung	33
3.1 Die Theorie der internen und externen Validität	33
3.1.1 Die Darstellung der Theorie der internen und externen Validität	33
3.1.2 Störfaktoren der internen und externen Validität	36
3.2 Die Darstellung der experimentellen Einzelfalluntersuchung	40
3.2.1 Die Logik der experimentellen Einzelfalluntersuchung	40
3.2.2 Die Arten (Variationen) der experimentellen Einzelfalluntersuchung	42
3.3 Die Anwendbarkeit der von Campbell und Stanley (1973) für die pädagogische Forschung vorgeschlagenen Quasi-Experimente	45
4. Wie kann die experimentelle Einzelfalluntersuchung die von Campbell & Stanley (1973) erwähnten acht Störfaktoren, die die interne Validität gefährden, ausschalten?	49
4.1 Die Ausschaltung der acht Störfaktoren der internen Validität	49
4.2 Ein weiterer Störfaktor, der die interne Validität möglicherweise gefährden kann	51
5. Wie kann die experimentelle Einzelfalluntersuchung	

externe Validität durch systematische Wiederholung erreichen?

- 3.5.1 Der Zusammenhang zwischen systematischer Wiederholung und externer Validität
- 3.5.2 Die direkte Wiederholung
- 3.5.3 Die systematische Wiederholung
- 4 Untersuchung II: Die Probleme in der Durchführung der experimentellen Einzelfalluntersuchungen
 - 4.1 Fragestellungen
 - 4.2 Methode
 - 4.2.1 Sammlung der Daten
 - 4.2.2 Auswahl der Zeitschriften
 - 4.2.3 Kriterien für die Aufnahme von Forschungsberichten
 - 4.3 Ergebnisse
 - 4.4 Zusammenfassung der Ergebnisse der Untersuchung II
 - 4.5 Diskussion der Anwendbarkeit der verschiedenen statistischen Prüfverfahren
- 5 Diskussion und Schlußfolgerungen: Wie kann die experimentelle Einzelfalluntersuchung für die pädagogische Forschung fruchtbar gemacht werden?
 - 5.1 Die Verwendung der experimentellen Einzelfalluntersuchung bei der Erforschung des sozialen Verhaltens
 - 5.2 Die Verwendung der experimentellen Einzelfalluntersuchung bei der Erforschung des Schulleistungsverhaltens
 - 5.3 Wie kann die experimentellen Einzelfalluntersuchung zur Theoriebildung herangezogen werden?
 - 5.3.1 Die Eigenschaften der Theorie
 - 5.3.2 Prozesse der Theoriebildung
 - 5.3.3 Die Bedeutung der experimentellen Einzelfalluntersuchung bei der Theoriebildung
- 6 Zusammenfassung
 - Literaturverzeichnis
 - Anhang

Einleitung

Die Akkumulation und Verarbeitung der präzisen Aussagen, die durch empirische Untersuchungen gewonnen werden, spielen eine große Rolle bei der Entstehung und Entwicklung der Wissenschaft. Die Wichtigkeit der empirischen Untersuchung in der Erziehungswissenschaft ist heutzutage auch wohl bekannt (vgl. Roth, 1964; Brezinka, 1964; Fischer, 1964). Im Prozeß der Verwissenschaftlichung der Pädagogik versucht man die experimentell-psychologischen Aussagen und ihre Methodologie in die Erziehungswissenschaft einzubürgern. Doch nach jahrzehntelanger Anstrengung, die wegen der Kluft zwischen der Theorie, die aus der psychologischen Laborforschung entsteht, und der Erziehungspraxis zu keinem befriedigenden Ergebnis führte, fängt man heute an, die Handlungsrelevanz der erziehungswissenschaftlichen Erkenntnisse zu erwägen (vgl. Krumm, 1979). Handlungsrelevanz bezieht sich, nach Krumms Ansicht, auf die Frage, "ob und wie ein gegebenes pädagogisches Problem gelöst werden kann" (Krumm, 1979, S.3). Dabei handelt sich um ein technologisches Problem. Es wird angenommen, daß das Forschungsergebnis nur dann für die Erziehungspraxis "handlungsrelevant" ist, wenn es in der erzieherischen Realität entstanden ist. Um die konzipierten erziehungswissenschaftlichen Erkenntnisse bzw. Handlungsalternativen zur Lösung des erzieherischen Problems kritisch zu prüfen, ist es unentbehrlich, eine für erzieherische Situation geeignete Untersuchungsmethodologie zu entwickeln. In diesem Fall sind für erziehungswissenschaftliche Experimente angemessene Versuchsanordnungen dringend gesucht (vgl. Hagemann, 1976). Eine solche für ein erziehungswissenschaftliches Experiment angemessene Versuchsanordnung soll nach König zwei Eigenschaften besitzen: die Wissenschaftlichkeit und die Praktikabilität, d.h. sie sollte zum einen so wissenschaftlich sein, daß sie eine intersubjektive Überprüfung der Individualtheorie ermöglicht, zum anderen muß sie so praktikabel sein, daß sie in der konkreten Situation anwendbar ist (vgl. König, 1978, S.135). Aber König fährt fort: "...Gerade in dieser doppelten Zielsetzung (möglichst intersubjektive Überprüfbarkeit und möglichst hohe Praktikabilität) dürfte die Hauptschwierigkeit bei der wissenschaftstheoretischen

Absicherung erziehungswissenschaftlicher Individualtheorie liegen"(König,1978,S.135).

Der Zweck der vorliegenden Arbeit liegt darin, die Anwendbarkeit von experimentellen Einzelfalluntersuchungen, die die Anforderungen der von König erwähnten doppelten Zielsetzung erfüllen könnten, für erziehungswissenschaftliche Experimente zu untersuchen.

1.1 Theorie-Praxis-Verhältnis

Die Pädagogik tendiert dazu, sich zu einer Erfahrungswissenschaft weiterzuentwickeln (vgl. Flehsig,1967). Die geisteswissenschaftliche Pädagogik wird wegen ihres Mangels an methodischem Vorgehen kritisiert. Werres (1975) sagt:"Sie (Geisteswissenschaftliche Pädagogik) versäumt es, ihre Theorie in einer verständlichen und begrifflich präzisen Sprache zu fassen und entzieht sich so der kritischen Diskussion: Theorie und Praxis leiden in gleicher Weise unter dieser Entwicklung: Die Vagheit und Vieldeutigkeit der Begriffe sind Anlaß zu endlosen und fruchtlosen Debatten. Die sich darin ausformende Wortmystik wird von der Praxis nicht mehr verstanden und kommt ihren Bedürfnissen kaum noch entgegen." (S.16).

Aber nach jahrzehntelanger erfahrungswissenschaftlicher Orientierung stellen die Erziehungswissenschaftler die Frage, ob die durch empirische Forschungen aufgebauten Theorien in der Lage sind, vor dem Hintergrund einer schwieriger und komplizierter werdenden erzieherischen Praxis konstruktive und produktive Hilfen zur Lösung zahlreicher und drängender Fragen zu vermitteln (vgl. Werres,1975, S.19). Man macht der reinen empirischen Forschung in der Erziehungswissenschaft den Vorwurf, daß sie unfähig ist, ihre Ergebnisse in der Erziehungspraxis anzuwenden. Holzkamp (1972) kritisiert an der Laboratoriumsforschung,"daß die Resultate der psychologischen Lernforschung etwa zur Verbesserung der Unterrichtsmethodik in der Schule, in der es schließlich auch um Lernen geht, kaum etwas Wesentliches beigetragen haben" (S.23).

Travers (1972) hat die Kluft zwischen der wissenschaftlichen Laborforschung und der Erziehungspraxis umrissen:

"Die Beziehung zwischen der Forschungsarbeit des Wissenschaftlers im Labor und der pädagogischen Praxis war lange kontrovers. Während es in den Naturwissenschaften gut entwickelte Verbindungen zwischen dem sogenannten Laboratoriumswissenschaftler und dem Hersteller nützlicher Produkte gibt, bestehen vergleichbare Verbindungen im Falle der Verhaltenswissenschaft nicht. Diese Wissenschaften haben keine Vertreter, die man mit jenen Technikern vergleichen kann, die Laborentdeckungen in Erzeugnisse und Dienstleistungen umwandeln, derer die Gesellschaft bedarf" (S.78).

Vielleicht kann man die reine wissenschaftliche Forschung so verteidigen, daß das Ziel der Wissenschaft die Theoriebildung oder Wahrheitssuche ist, daß die Erkenntnis ihren eigenen Gehalt hat, weshalb der Wissenschaftler sich nicht zu sorgen braucht, ob seine Theorie der Praxis nützlich ist oder nicht.

Wegen dieser Argumentation stellt sich die Frage: soll die empirische Erziehungswissenschaft nur der erziehungswissenschaftlichen Theoriebildung dienen oder auch der Erziehungspraxis?

Lochner (1963) ist ein Verteidiger der wertfreien Erziehungswissenschaft. Seiner Meinung nach ist die Erziehungswissenschaft diejenige theoretische, selbständige, reine Wissenschaft, die sich auf die Gesamtheit des Erziehungsphänomens richtet (vgl. Lochner, 1963, S.415). Er behauptet, daß Wissenschaft außerhalb des Wahrheitswertes keinerlei gültige Werte aufstellen kann. Auf die Frage, was der Mensch tun soll, kann Wissenschaft keine Antwort geben. Sie muß sich daher bemühen, wertfrei zu bleiben. Wissenschaft kann gegenüber den verschiedenen Wertordnungen, die in der Welt in unlöslichem Kampf untereinander stehen, nicht Partei ergreifen; soll es eine Erziehungswissenschaft als Wissenschaft geben, dann muß sie sich an dieses Prinzip der Wertfreiheit halten (Lochner, 1963, S.432). Er schreibt: "Allgemein gültige Sätze kann man nur im Raum einer wertfreien Wissenschaft formulieren; in ihnen kann nur ausgesagt werden, was geschieht, nicht aber, was geschehen soll." (S.434). Auch nach Kerlinger (1972) ist das grundlegende Ziel der Wissenschaft die Theoriebildung (vgl. Kerlinger, 1972, S.10-11).

Andere Erziehungswissenschaftler fordern, daß die erziehungswissenschaftliche Forschung auch der Erziehungspraxis dienen soll.

Roth & Friedrich (1975) betonen, daß von der Forschung oft konkrete Hilfe für die Problem- und Aufgabenbewältigung erwartet wird. Nach Werres soll der Prozeß der Erkenntnisgewinnung zwar in seiner formalen Struktur möglichst unabhängig von unmittelbaren politischen oder praktischen Einflüssen bleiben; aber dennoch sind Wissenschaft und Praxis in einem Rückkoppelungsprozeß interaktiv verbunden (Werres, 1975, S.53). Seines Erachtens hat eine Untersuchung Doppelzügigkeit, d.h. die empirische erziehungswissenschaftliche Forschung soll sowohl der Theoriebildung als auch der Erziehungspraxis dienen. Er sagt: "Die kurzen Überlegungen weisen auf eine Doppelzügigkeit der Untersuchung hin: Einmal müßte sie der Frage nachgehen, welche Konsequenzen die Anwendung analytischer Methoden für den Prozeß der Theoriebildung in der Erziehung nach sich zieht, d.h. wie sich ihr spezifischer Beitrag zur Entwicklung eines organisierten Wissens zum Gegenstandsbereich Erziehung ausweist; als Wissenschaft, die menschliche Handlungen zu ihrem Objektbereich hat, muß die vorliegende Untersuchung zum anderen ihr Augenmerk ebenso auf das Problem der Relevanz der empirischen Erziehungswissenschaft für die Praxis richten. Gerade die Klärung der letzten Frage scheint mir besonders wichtig zu sein, da die Praxis heute nachhaltig sachlich fundierte Hilfe bei der Lösung schwieriger Probleme von der Theorie erwartet." (Werres, 1975, S.19-20).

Flehsig versucht die Kluft zwischen Theorie und Praxis durch technologische Experimente zu überbrücken (vgl. Flehsig, 1967). Er schreibt: "Nun ist es im Bereich der Unterrichtsforschung in der Regel noch ein weiter Weg vom experimentell gesicherten Einzelbefund zur wirksamen Verbesserung der Unterrichtspraxis. Es gilt nämlich, die unter isolierten (Labor-) Bedingungen gewonnenen Einsichten in Prozesse des Lehrens und Lernens zu integrieren mit jenen Faktoren, welche üblicherweise die Unterrichtrealität bestimmen. Eine bloße Ableitung didaktischer Maßnahmen aus den Befunden der 'reinen' Forschung verbietet sich von daher. Die Anwendung wird im

allg. so vor sich gehen, daß weitere Informationen und zusätzliche Relevanzkriterien hinzutreten müssen, um eine neue Hypothese bilden zu können. Diese würden sich dann auf den Bedingungs-zusammenhang beziehen, von dem die Anwendung des im Forschungsexperiment gesicherten Befundes in einer begrenzten Unterrichtspraxis abhängig sein wird. Sodann wird — wie im Fall des Laborexperiments — auch diese Hypothese der experimentellen Bestätigung unterworfen. Experimente dieser Art sollen als technologische Experimente bezeichnet werden." (Flehsig, 1967, S.404).

Nach Flehsig besteht die Hauptfunktion der technologischen Experimente darin, zu überprüfen, inwieweit die unter Laborbedingungen festgestellten Gesetzmäßigkeiten auch in der Realsituation erhalten bleiben, bzw. in welcher Weise sie verändert werden (Flehsig, 1967, S.411). Brezinka ist ebenfalls der Meinung, daß im Hinblick auf die in der Erziehungspraxis zu lösenden Probleme die Erziehungswissenschaft in erster Linie eine technologische Wissenschaft ist (Brezinka, 1971, S.32).

An dieser Stelle stellt sich die Frage, warum man die in der 'reinen' Forschung gewonnenen Befunde nicht unmittelbar in der Praxis verwerten kann, sondern das technologische Experiment als Vermittler zwischen Theorie und Praxis braucht.

Travers hat eine diese Frage betreffende Untersuchung beschrieben: "Ein von der Regierung (USA) unterstützter Forschungsbericht jüngsten Datums von R.R. Mackie und P.R. Christensen (1967) hat versucht, das Problem abzutasten, wie man Möglichkeiten der engeren Beziehung zwischen der Laborforschung in Psychologie und der Führung praktischer Institutionen, z.B. der Bildung, entwickeln könnte. Nach Mackie und Christensen besteht die Hauptschwierigkeit der Anwendung von Laborforschung auf Situationen außerhalb des Labors darin, daß die im Labor geltenden Variablen unter den Bedingungen des wirklichen Lebens nicht ohne weiteres identifizierbar sind. Im Falle der Naturwissenschaft, so scheint es, existiert das Problem nicht in gleichem Maße. Die Wirkungen der Gravitation lassen sich in vielen praktischen Situationen feststellen, genauso wie der Dehnungseffekt einer Substanz unter Temperaturerhöhung. Im Gegensatz dazu betrachtet man,

wie schwer feststellbar die einzelnen Verstärkungsmomente des Verhaltens sind, das sich im Klassenzimmer abspielt, gegenüber den Verstärkungsmomenten im Laborversuch." (Travers, 1972, S.79).

Anders als die chemischen oder physikalischen Substanzen, die relativ homogene Eigenschaft haben und von der Umwelt unter bestimmter Konstanz gleicherweise beeinflußt werden, haben die Versuchspersonen in der erziehungswissenschaftlichen Forschung unterschiedliche Bedürfnisse, Lerngeschwindigkeit, Beharrlichkeit, Vorkenntnisse, usw. und können von der Umwelt unterschiedlich beeinflußt werden. Deshalb ist es schwer, die in der kontrollierten Laborsituation festgestellten Gesetzmäßigkeiten auch in den natürlichen Schulsituationen ohne weiteres zu erwarten. Es ist sehr wahrscheinlich, daß die Laborsituation und die Realsituation unterschiedliche Einflüsse auf die Verhaltensweisen der Versuchspersonen ausüben.

Holzkamp stellt fest, " daß die isolierten, reduzierten, labilisierten Bedingungen der experimentellen Untersuchungen in der Unterrichtssituation kaum vorliegen, so daß nicht entschieden werden kann, ob die im Experiment aufgewiesenen Effekte in der Schulpraxis eine berücksichtigenswerte Größenordnung besitzen." (Holzkamp, 1972, S.24). Bracht & Glass (1968) sind der Ansicht, daß die pädagogischen Laborexperimente keine Basis für die Verallgemeinerung auf eine Vielzahl von Situationen bieten, in denen der Mensch normalerweise mit seiner Umwelt interagiert. Snow (1974) vermutet, daß die größte Gefahr für die externe Validität zum Vorschein kommen wird, wenn das Experiment der Natur des zu untersuchenden Verhaltens nicht entspricht.

Es ist klar geworden, daß viele Bedingungen, die in der natürlichen Situation vorliegen, in der Laborforschung reduziert oder geändert werden. Dies ist erforderlich, um im Labor eine kontrollierbare Situation zu schaffen, so daß die innere Validität des Experiments abgesichert werden kann. "Wir müssen alle experimentellen Kontrollen berücksichtigen, um auszuschließen, daß das Experiment durch andere Störfaktoren zu dem Ergebnis gelangte. D.h., alle Randbedingungen, die ein Ergebnis des Versuches beeinflussen

können, müssen wirklich genügend kontrolliert werden, ... dann kann sich die Schlußfolgerungen aus den Ergebnissen rechtfertigen lassen." (Edwards, 1971, S. 21-22).

Zur Überbrückung der Kluft zwischen Laborexperiment und Praxis schlägt Flechsig das technologische Experiment vor, um die in Laborexperimenten erforschten Gesetzmäßigkeiten durch Feldexperimente in der Praxis zu überprüfen (vgl. Flechsig, 1967, S. 410).

Es taucht das Problem auf, wie man beim Feldexperiment den Einfluß der Störfaktoren, die die interne Validität gefährden können, ausschalten oder kontrollieren kann. Oder darf man auf die interne Validität des Experiments verzichten?

Pädagogen, die die Handlungsforschung befürworten, neigen dazu, auf die für die wissenschaftliche Arbeit unentbehrliche interne Validität zu verzichten. Haerberlin hat dies eindeutig beschrieben: "Prinzipien, die für den Objektivitätsbegriff empirisch-analytischer Forschung unabdinglich sind, wie Kontrolliertheit und Wertfreiheit, werden geradezu in ihrer Umkehrung zu Prinzipien der Handlungsforschung. Kontrolliertheit im Sinn von Unveränderbarkeit des Forschungsfelds und Wertfreiheit im Sinn von Unparteilichkeit des Forschers während des Forschungsprozesses würden die Handlungsforschung am Erreichen ihres Zieles, nämlich verändernde und parteiische Einflußnahme auf das Forschungsfeld, hindern." (Haerberlin, 1975, S. 664). Heinze et al. (1975) vertreten den Standpunkt: "Handlungsorientierte Evaluation dürfe nicht in das methodische Korsett standardisierter und kontrollierter Untersuchungsansätze gezwängt werden." (S. 211).

Aber eine Untersuchung, die nicht mit der wissenschaftlichen Objektivität durchgeführt wurde, ist nicht wiederholbar, und die Befunde aus einer solchen Untersuchung können nicht anerkannt werden. Das technologische Experiment soll nach Flechsig an erfahrungswissenschaftliche Prinzipien gebunden werden: "Je stärker die Bindung dieser Technologie an erfahrungswissenschaftliche Prinzipien ist, desto weniger gerät sie in die Gefahr, modischen Aberrationen zu verfallen." (Flechsig, 1967, S. 412). Eine bestimmte (explizite) und systematische Forschungsmethode ist nach Reif (1972) die

wesentliche Voraussetzung für den Fortschritt der angewandten Erziehungswissenschaft. "In der Tat ist es besser, erfolglos aber bestimmt zu sein als erfolgreich aber unbestimmt. Im ersten Fall hat man immerhin eine gute Chance, seine Schwierigkeit zu diagnostizieren und sie zu überwinden. Aber im zweiten Fall weiß man nicht, warum er erfolgreich ist, und deshalb hat er nichts gelernt, um seine künftigen Erfolge oder Verbesserungen zu sichern." (Reif, 1978, S. 3. übersetzt vom Verfasser). Um die Wissenschaftlichkeit einer Untersuchung zu gewährleisten, darf also auf die Kontrolle der internen Validität nicht verzichtet werden.

Zur Durchführung eines technologischen Experiments bedarf es deshalb einer Forschungsmethode, die einerseits in der realen Schulsituation anwendbar ist und andererseits den Einfluß der Störfaktoren auf die Ergebnisse des Experiments beseitigen kann. Flehsig selbst hat diesbezüglich keine Vorschläge gemacht.

1.2 Welche Forschungsstrategie ist für praxisorientierte Experimente angemessen?

In der Literatur werden einige Forschungsmethoden im Hinblick auf ihre Anwendbarkeit für praxisorientierte Experimente diskutiert:

a) ATI-Forschung

In ihrem Buch "Aptitudes and instructional methods: A handbook of research on interactions" haben Cronbach & Snow (1977) die Methodologie der ATI-(Aptitude-Treatment Interaction) Forschung ausführlich dargestellt. Bei dieser Methode versucht man mit dem Kontrollgruppendesign (Cronbach, 1975b, S. 116) oder faktoriellen Design (Bracht, 1975, S. 95) die Untersuchung in der natürlichen Situation durchzuführen. Mit Varianz-, Korrelations- oder Regressionsanalysen werden die Wechselwirkungen zwischen Schülermerkmalen und Unterrichtsmethoden analysiert. Wird eine ATI festgestellt, so bedeutet das, daß eine bestimmte Unterrichtsmethode bei einem bestimmten Schülermerkmal wirksamer ist als bei einem anderen. Man versucht, auf der Grundlage der Ergebnisse der ATI-Forschung die Unterrichtsmethoden an individuelle Unterschiede anzupassen (Cronbach, 1975a).

Aber bis heute sind die Ergebnisse der ATI-Forschung nicht sehr eindrucksvoll. Cronbach selbst sagt: "Es ist nicht selten der Fall, daß zwei oder mehr Untersuchungen, die über ungefähr dieselben Merkmals- und Methodenvariablen verfügen, sich widersprechen." (Cronbach, 1975a, S. 51).

Es gibt zwei mögliche Interpretationen für diese Unbeständigkeit der Ergebnisse aus ATI-Forschungen:

- (1) Das Treatment wird von Störfaktoren beeinflusst. Wenn man mit dem traditionellen Kontrollgruppen-Experimental-design die Untersuchung in der natürlichen Situation durchführt, verletzt man die Voraussetzung des Experimentaldesigns — die experimentelle Kontrolle (s. Edwards, 1971, S. 21-22), besonders, wenn die Untersuchung sich über Wochen, Monate, Semester oder sogar Jahre erstreckt. Die Forscher der ATI neigen dazu, ökologische Repräsentativität der Untersuchung auf Kosten der inneren Validität zu erzielen. "Zeitliche Ausdehnung eines Treatments macht die fremden Einflüsse auf die experimentellen Effekte möglich, ... Aber die sogenannten fremden Einflüsse sind zuträglich im Sinn des repräsentativen Designs, wenn sie den experimentellen Treatments ermöglichen, der natürlichen Situation des schulischen Lernens zu ähneln. Kontrolle wird zur Zeit als weniger wichtig als die Repräsentativität betrachtet." (Snow, 1974, S. 281, übersetzt vom Verfasser). Deshalb ist es sehr wahrscheinlich, daß die Einflüsse der Störungen durch die unkontrollierten Störfaktoren für die inkonsistenten und nicht-wiederholbaren Ergebnisse der ATI-Forschung verantwortlich sind.
- (2) Die Unbeständigkeiten (inconsistencies) werden in Zukunft mit Interaktionen höheren Grades interpretiert werden können (Cronbach, 1975b, S. 119). Aber je mehr Variablen man in der Untersuchung manipuliert, desto schwerer kann man die Effekte der Einzelvariablen feststellen. Vielleicht kann man dies durch die Verbesserung der Methodologie erreichen. Snow (1974) hofft: "Obwohl viele der gegenwärtig vorhandenen Befunde wahrscheinlich nicht stabil und nicht wiederholbar sind, kann das Streben der ATI durch die verbesserte Hypothese und Methodologie endlich die Beschaffenheit von Aptitude für Lernen entdecken und liefert einen neuen Schlüssel für die Verbesserung des Unterrichtsverfahrens."

(S.271, übersetzt vom Verfasser).

Ob die ATI-Forschung für die technologischen Experimente angemessen ist, kann heute noch nicht beurteilt werden.

Schulman (1975) hat auch vier Arten von Forschungsstrategien für das Feldexperiment vorgeschlagen: 1) epidemiologische Strategien, 2) die Entdeckung von grammatischen Strukturen im Verhalten, 3) Simulation in der Forschung, und 4) multivariate experimentelle Längsschnittstudien (S.7-11). Die ersten zwei Strategien sind medizinischer und linguistischer Forschung entlehnt worden. Sie können nur der Hypothesenformulierung dienen und als explorative Experimente im Flechsigschen Modell betrachtet werden. Die dritte Strategie versucht, künstliche Umwelt im Labor zu schaffen, die der tatsächlichen Umwelt so ähnlich wie möglich ist. Diese Strategie muß trotzdem als Laborexperiment betrachtet werden, weil fraglich ist, ob die natürliche Situation im Labor simuliert werden kann. "Der Versuch, unterrichtliche Vorgänge im Laborexperiment rekonstruierbar zu simulieren, bedingt aus forschungsmethodologischen Gründen vielfach Reduktionen der unterrichtlichen Wirklichkeit, die den systematischen Rückbezug der Ergebnisse auf die Unterrichtsrealität und damit auch die Planung und Durchführung realitätsnaher technologischer Experimente ungeheuer erschweren, wenn nicht gar ausschließen." (Hagemann, 1976, S.132). Die vierte Strategie kann der ATI zugeordnet werden.

Campbell & Stanley (1973) haben das Quasi-Experimentaldesign auch für Feldforschung vorgeschlagen. Solche Quasi-Experimentaldesigns werden später eingehend zusammen mit der experimentellen Einzelfalluntersuchung erörtert.

b) Experimentelle Einzelfalluntersuchung (heißt auch Einzelfall-Experimentaldesign)

Die experimentelle Einzelfalluntersuchung findet in der Verhaltensmodifikation Anwendung. Im Gegensatz zum traditionellen Experimentaldesign werden in ihm Baseline- und Behandlungsphasen abgewechselt. In der Baseline-Phase wird das zu untersuchende Verhalten der Versuchsperson in der natürlichen Situation ohne Treatment beobachtet. In dieser Phase ist das Verhalten der Versuchsperson für die Einflüsse aller unbekannten Störfaktoren offen, man informiert sich,

wie die Verhaltenskurve ohne Treatment aussieht. Wenn der Verlauf der Kurve stabil ist, führt man das Treatment ein. Auch das Treatment wird so lange durchgeführt, bis die Kurve stabil ist. Aufgrund des Unterschieds zwischen den Kurven der Baseline- und der Treatmentphase kann man beurteilen, ob das Treatment einen Effekt hat oder nicht. Um zu sichern, daß das Treatment für die Veränderung (den Effekt) allein verantwortlich ist, nimmt man in einer zweiten Baseline-Phase das Treatment zurück. Wenn sich die Kurve der zweiten Baseline-Phase in Richtung der ersten Baseline-Phase ändert, ist der Effekt des Treatments feststellbar. Hersen & Barlow haben in ihrem Buch "Single case experimental designs" Entwicklung, Methodologie und Arten der experimentellen Einzelfalluntersuchung skizziert. Sie haben auch die Streitpunkte und die mögliche Koordination zwischen Gruppen- und Einzelfalldesign diskutiert (Kap.2). Die experimentelle Einzelfalluntersuchung ist in der natürlichen Situation durchführbar und durchgeführt worden. In dem seit 1968 erscheinenden Journal of Applied Behavior Analysis sind überwiegend Untersuchungen mit Einzelfall-Experimentaldesign dargestellt.

Ist die experimentelle Einzelfalluntersuchung eine angemessene Forschungsmethode für die technologischen Experimente? Folgende Gründe unterstützen ihre Angemessenheit:

1) Obwohl es sehr unwahrscheinlich ist, jedem einzelnen Schüler Privatunterricht zu erteilen (vgl. Bloom, 1970, S.18), soll hinsichtlich der pädagogischen Verantwortung das einzelne Kind nicht vernachlässigt werden, wenn es in der Schule Störverhalten zeigt oder nicht von dem Unterricht profitieren kann. Um das Problem eines einzelnen Schülers zu lösen, ist die experimentelle Einzelfalluntersuchung angebracht. Die ATI-Forschung versucht, den Unterricht dem Individuum anzupassen, aber sie gibt nicht an, welchem Treatment ein Schüler mit bestimmten Aptitudes zugewiesen werden soll (vgl. Cronbach & Snow, 1977, S.520). Krumm (1977) schreibt: "Wir kennen alle aus der Kontrollgruppenforschung genügend Beispiele bzw. Experimente, in denen sich die Mittelwerte zwar in erwünschter Weise verschoben, aber doch einige einzelne Versuchspersonen nach dem Treatment absolut schlechter

abschnitten als vor dem Treatment,..."(S.6). Ob ein Treatment das Verhalten des einzelnen Schülers in erwünschter Weise ändern kann, muß mit der experimentellen Einzelfalluntersuchung untersucht werden.

2) Hinsichtlich der theoretischen Überlegung ist die experimentelle Einzelfalluntersuchung auch gerechtfertigt. Die experimentelle Einzelfalluntersuchung soll durch intensive Untersuchung bei einer einzelnen Versuchsperson den Mechanismus der Verhaltensänderung entdecken. So behauptet Reif in seinem Artikel:"Toward an Applied Science of Education: Some key questions and directions": " Obwohl viele statistische Daten in der Erziehung gesammelt wurden und eventuell benutzt werden können, ist eine solch grobe statistische Information meist nicht geeignet, Einsicht in die zugrundeliegenden Mechanismen oder in die Vorgänge der Verbesserung der erzieherischen Modelle oder Methoden zu geben. Zu diesem entscheidend wichtigen Zweck ist es notwendig, ausführliche Informationen aus den sorgfältigen Beobachtungen von den kontrollierten Experimenten mit individuellen Schülern zu erfahren. (Eine Analogie besteht darin, daß die Entwicklung eines wirkungsvollen Treatments für eine Krankheit wesentlich von den eingehenden biochemischen Forschungen abhängt, die die grundlegenden Mechanismen erläutern und zum Design der spezifischen Interventionsformen führen. Die Sammlung der statistischen Daten spielt meistens eine sehr periphere Rolle)."(Reif,1978,S.3, übersetzt vom Verfasser)

Die experimentelle Einzelfalluntersuchung scheint eine angemessene Grundlage für das erziehungswissenschaftliche technologische Experiment zu sein. Aber bisher gibt es noch keine Untersuchung darüber, ob und wie die experimentelle Einzelfalluntersuchung in der erziehungswissenschaftlichen Forschung fruchtbar gemacht werden kann. Die vorliegende Arbeit versucht, diese Frage zu lösen. Es werden zwei empirische Untersuchungen durchgeführt: die erste gibt einen Überblick über die gegenwärtigen Forschungslage der Erziehungswissenschaft in der Bundesrepublik. In dieser Untersuchung wird auch analysiert, inwieweit die experimentelle Einzelfalluntersuchung in der erziehungswissenschaftlichen

Forschung angewendet worden ist. Die zweite Untersuchung behandelt die Validität der experimentellen Einzelfalluntersuchung. Schließlich werden die potentiellen Beiträge der experimentellen Einzelfalluntersuchung zur erziehungswissenschaftlichen Forschung erörtert.

2 Untersuchung I: Die gegenwärtige erziehungswissenschaftliche Forschungslage in Deutschland

2.1 Fragestellung

"Wenn das Ziel pädagogischer Forschung die Entwicklung einer in sich konsistenten und in der Praxis brauchbaren Theorie ist, so sind die heutigen Forscher von diesem Ziel so weit entfernt wie davon, das Wetter zu bestimmen (Schulman, 1974, S. 798). Sind die heutigen Forscher noch in der Phase der Grundlagenforschung oder haben sie schon versucht, die in der Grundlagenforschung erworbenen Ergebnisse in die Praxis umzusetzen? Wenn die heutigen erziehungswissenschaftlichen Forscher noch nicht über die erste Phase hinausgekommen sind, dann ist die Anzahl der in den nicht-natürlichen Situationen durchgeführten Untersuchungen viel größer als die, die in den natürlichen Situationen durchgeführt wurden. Die Anzahl der Untersuchungen wird durch eine Stichprobe aus den Forschungsberichten in den deutschen erziehungswissenschaftlichen Zeitschriften repräsentiert.

Nullhypothese: Es gibt keinen Unterschied zwischen der Anzahl der in natürlichen und nicht-natürlichen Situationen durch Forschungsberichte dokumentierten Untersuchungen.

Operationaldefinition:

- (a) Die natürliche Situation: die Situation, in der die Alltagsbedingungen (Lehrer, Lehrmethode, Lehrstoff, Organisation des Klassenzimmers hinsichtlich der Schüler und/oder physikalischer Umwelt, Stundenplan) nicht geändert oder reduziert wurden, abgesehen davon, daß sie als unabhängige Variablen manipuliert wurden.
- (b) Die nicht-natürliche Situation: die Situation, in der die Alltagsbedingungen total oder teilweise geändert oder reduziert wurden, abgesehen davon, daß sie als unabhängige Variablen manipuliert wurden.

2.2 Methode

2.2.1 Auswahl der Zeitschriften

Vier deutschsprachige pädagogische Zeitschriften werden

für die vorliegende Untersuchung ausgewählt:

- 1) Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie (ZEPP)
- 2) Psychologie in Erziehung und Unterricht (PEU)
- 3) Zeitschrift für erziehungswissenschaftliche Forschung (ZeF)
- 4) Die Deutsche Schule (DS)

Die zu untersuchenden Jahrgänge jeder Zeitschrift gehen von 1972 bis 1976.

Folgende Überlegungen haben zu dieser Auswahl geführt:

- (a) Die auszuwählenden Zeitschriften sollen überwiegend empirische erziehungswissenschaftliche Forschungsberichte enthalten; um diesem Kriterium zu entsprechen, konnte auf eine Literaturübersicht zurückgegriffen werden.

Nach der Veröffentlichung der erziehungswissenschaftlichen Schwerpunktprogramme für die Förderung der erziehungswissenschaftlichen Forschung (DFG-Kommission Erziehungswissenschaft, 1974) analysierten Heidenreich und Heymann die Lehr-Lern-Forschungspunkte aufweisen den deutsch- und englischsprachigen pädagogischen Forschungsberichte nach 10 Prinzipien der Lehr-Lern-Forschung. Im Rahmen einer extensiven Literatursichtung wählten sie die für die Lehr-Lern-Forschung relevanten Monographien, Zeitschriftenartikel und Bücher aus und ordneten diese tabellarisch nach thematischen Schwerpunkten an (vgl. Heidenreich & Heymann, 1976, S.242, Tabelle 1: Literaturübersicht). Die unten aufgelisteten deutschen Zeitschriften wurden in die vorliegende Untersuchung aufgenommen und als eine Stichprobe von Zeitschriften, die empirische erziehungswissenschaftliche Forschungsberichte enthalten, angesehen. Es handelt sich um folgende Zeitschriften:

- 1) Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie (20)
- 2) Psychologie in Erziehung und Unterricht (32)
- 3) Die Deutsche Schule (3)
- 4) Zeitschrift für erziehungswissenschaftliche Forschung (3)
- 5) Zeitschrift für experimentelle angewandte Psychologie (3)
- 6) Unterrichtswissenschaft (3)

7) Deutsche Berufs- und Fachschule (3)

8) Gruppendynamik (3)

9) Bildung und Erziehung (2)

Von diesen neun Zeitschriften wurden vier Zeitschriften nach der Anzahl der in ihnen enthaltenen empirischen Forschungsberichte ausgewählt. Die Anzahl der empirischen Forschungsberichte jeder Zeitschrift wurde anhand einer Stichprobe für den Jahrgang 1976 geschätzt. Hinter den Titeln ist die Anzahl der im Jahrgang 1976 enthaltenen Forschungsberichte angegeben. Empirische Forschungsberichte wurden operational definiert als Forschungsberichte, die eindeutig über die Durchführung des Experiments und die quantitativen Ergebnisse berichten. Da die ersten beiden Zeitschriften (Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie und Psychologie in Erziehung und Unterricht) jeweils 20 und 32 empirische Forschungsberichte enthalten, werden sie bedenkenlos aufgenommen. Weil die anderen Zeitschriften ungefähr gleich viele empirische Forschungsberichte haben, werden die dritte und vierte Zeitschrift (Die Deutsche Schule und Zeitschrift für Erziehungswissenschaftliche Forschung) willkürlich aufgenommen.

- (b) Auch Jahrgänge (1972-1976) sind willkürlich bestimmt worden und werden als eine Stichprobe der "gegenwärtigen Forschungslage" betrachtet.

2.2.2 Auswahl der Artikel

Die vorliegende Arbeit interessiert sich nur für die Anwendbarkeit der experimentellen Forschungsanordnungen in der erziehungswissenschaftlichen Forschung, besonders in der Unterrichtsforschung. Deshalb werden nur solche Forschungsberichte ausgewählt, denen experimentelle Designs zugrunde liegen. Die traditionellen Experimentaldesigns, die die interne Validität beinhaltenden Quasi-Experimentaldesigns (d.h. die Anordnungen 8,9,13a, und 14 bei Campbell & Stanley, 1973) und die Einzelfall-Experimentaldesigns werden als experimentelle Forschungsanordnungen betrachtet. Die vor-experimentellen Designs und die keine vollständige

interne Validität beinhaltenden Quasi-Experimentaldesigns bei Campbell & Stanley (1973), A-B-Design bei Einzelfall-Experimentaldesigns und die Korrelationsdesigns ohne Treatment werden ausgeschlossen.

2.2.3 Sammlung der Daten

Nach den folgenden Items werden Informationen aus den ausgewählten Forschungsberichten gesammelt:

- (a) Titel
- (b) Unabhängige Variablen
- (c) Abhängige Variablen
- (d) Forschungsanordnung
- (e) Wurde die Untersuchung in der natürlichen oder in der nicht-natürlichen Situation durchgeführt?

2.3 Ergebnisse und Diskussion

Insgesamt wurden 77 Artikel, die über experimentelle Untersuchungen berichten, als Stichprobe aus den vier ausgewählten erziehungswissenschaftlichen Zeitschriften gezogen. Die Prozentsätze der ausgewählten Artikel jeder Zeitschrift bei den Jahrgängen 1972-1976 werden in Tabelle 2-1 aufgeführt. Die Gesamtzahl der Artikel jedes Jahrgangs ist nach den Angaben des Inhaltsverzeichnisses jedes Jahrgangsbandes gezählt worden (die Zeitschriftenreferate, Notizen, Buchanzeigen, Buchbesprechungen, Zusammenfassungen ausgewählter Arbeiten aus anderen Zeitschriften wurden nicht berücksichtigt). Viele Artikel, die über empirische aber nicht experimentelle Forschung berichten, sind nicht in die Stichprobe aufgenommen worden, da die unabhängigen Variablen nicht manipuliert wurden. Dies ist z.B. der Fall, wenn ein bestimmter Test bei verschiedenen Altersgruppen durchgeführt wurde, um Unterschiede zwischen den Altersgruppen zu erfassen, aber keine Behandlung eingesetzt wurde. Deswegen ist die Anzahl der ausgewählten experimentellen Artikel geringer als die am Anfang zur Auswahl der Zeitschriften gezählten empirischen Artikel.

Zur Überprüfung der Nullhypothese (Es gibt keinen Unterschied zwischen der Anzahl der in natürlichen Situationen und der Anzahl der in reduzierten Situationen durchgeführten

Untersuchungen) wird die Artikelanzahl im Hinblick auf drei Variablen (Untersuchungssituation, Versuchsanordnung und Zeitschrift) in Tabelle 2-2 zusammengestellt.

Im allgemeinen wurden die Situationen als "reduziert" betrachtet, wenn

1. die Versuchszeit weniger als eine Unterrichtsstunde, z.B. nur 10 Min. betrug,
2. die Vpn einzeln im isolierten Versuchsraum untersucht oder getestet wurden,
3. die Versuchsgruppen neu zusammengestellt und unterrichtliche Freistellung der Kinder für das Experiment vorgevorgenommen wurden,
4. die gebrauchten Reizmaterialien sinnleere Silben waren,
5. die Situation simuliert war.

Im Gegensatz dazu wurden die Situationen als "natürlich" bezeichnet, wenn

1. die Unterrichtssituation der Vpn während des gesamten Versuchszeitraumes völlig normal blieb, keine der zur Ausschaltung bzw. Kontrolle von Störvariablen praktizierten Maßnahmen den schulischen Alltag verfälschten,
2. die Untersuchung sich auf einen längeren Zeitraum als eine Unterrichtsstunde (z.B. drei Wochen oder sogar vier Semester) erstreckte,
3. die Inhalte der Schulstunden, das Arbeitsmaterial und der Ablauf des Experiments von den Lehrern bestimmt wurden und dem Unterrichtsplan entsprachen,
4. die Situation echt war, z.B. die Behandlung für Prüfungsangst kurz vor einer mündlichen Prüfung stattfand oder die Untersuchung im Jugendheim oder im Hotel durchgeführt wurde.

Es sind 27 Artikel, die über die in den reduzierten Situationen durchgeführten Untersuchungen berichten, während bei 51 Artikeln, die Autoren die Untersuchungen in den natürlichen Situationen durchführten. Die Nullhypothese muß zurückgewiesen werden ($\chi^2=7.38$, $Fg=1$, $p<0.01$). Dieses Ergebnis besagt, daß die gegenwärtigen erziehungswissenschaftlichen experimentellen Forschungen in der Bundesrepublik viel mehr an der Feldforschung als an der Laborforschung orientiert sind. Aber diese Aussage darf nur vorsichtig

Tabelle 2-1: Prozentsätze der ausgewählten Artikel bei den vier ausgewählten erziehungswissenschaftlichen Zeitschriften ¹⁾

Zeit- schriften	Artikel	Jahrgänge					Σ
		'72	'73	'74	'75	'76	
<u>ZEPP</u>	gesamt	23	22	21	24	29	119
	ausgewählt	5	8	5	2	9	29
	%	21.7	36.4	23.8	8.3	31.0	24.4
<u>PEU</u>	gesamt	49	37	45	45	41	217
	ausgewählt	8	8	7	3	14	40
	%	16.3	21.6	15.6	6.7	34.1	18.4
<u>ZeF</u>	gesamt	13	19	12	13	16	73
	ausgewählt	0	0	1	1	0	2
	%	0.0	0.0	8.3	7.7	0.0	2.7
<u>DS</u>	gesamt	66	77	67	65	61	336
	ausgewählt	3	1	1	0	1	6
	%	4.5	1.3	1.5	0.0	1.6	1.8
Σ	gesamt	155	155	145	147	147	745
	ausgewählt	16	17	14	6	24	77
	%	10.6	11.0	9.7	4.1	16.3	10.3

ZEPP = Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie

PEU = Psychologie in Erziehung und Unterricht

ZeF = Zeitschrift für erziehungswissenschaftliche Forschung

DS = Die Deutsche Schule

% = ausgewählte Artikel
Gesamtzahl der Artikel

1) Die Seitenangabe der ausgewählten Artikel siehe Anhang 1.

Tabelle 2-2:Häufigkeitsverteilung hinsichtlich der verschiedenen Situationen, Versuchsanordnungen und Zeitschriften

Situationen	Versuchsanordnungen	Zeitschriften					Σ	$\Sigma\Sigma$
		ZEPP	PEU	ZeF	DS			
reduziert	Kontrollgruppen-							
	design	4	3	0	1	8		
	faktorielles Design	14	2	0	0	16		
	Einzelfall-							
	Experimentaldesign	0	0	0	0	0		
	randomisiertes						27	
natürlich	Ausbalancierungsdesign	1	0	0	0	1		
	Quasi-Experiment	1	1	0	0	2		
	Kontrollgruppen-							
	design	8	22	1	4	35		
	faktorielles Design	2	5	1	0	8		
	Einzelfall-							
	Experimentaldesign	0	3	0	0	3		
	randomisiertes						51	
	Ausbalancierungsdesign	0	3	0	1	4		
	Quasi-Experiment	0	1	0	0	1		

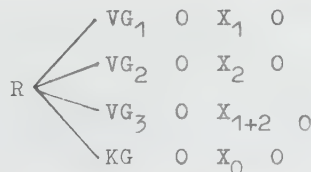
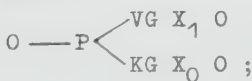
interpretiert werden, weil der Häufigkeitsunterschied zwischen den reduzierten und den natürlichen Situationen im Zusammenhang mit der Art der Zeitschrift steht. Die Tabelle 2-2 kann auf eine kleinere Tabelle mit zwei Variablen (Situation und Zeitschrift) reduziert werden:

Situation	Zeitschrift				
	<u>ZEPP</u>	<u>PEU</u>	<u>ZeF</u>	<u>DS</u>	
reduziert	20	6	0	1	27
natürlich	9	34	2	5	50
	29	40	2	6	77

In der reduzierten Tabelle sind nur 50 Artikel über Untersuchungen in der natürlichen Situation und damit einer weniger als in Tabelle 2-2 aufgeführt. Das kommt dadurch zustande, daß ein Artikel in ZEPP über zwei Untersuchungen mit zwei verschiedenen Versuchsanordnungen berichtet. Die reduzierte Tabelle zeigt, daß die drei Zeitschriften PEU, ZeF und DS über mehr Feldforschungen als Laborforschungen berichten, dieses Verhältnis bei ZEPP aber umgekehrt ist. Eine Überprüfung mit dem Chi-Quadrat-Test bestätigt diesen Unterschied ($\chi^2=18.03$, $Fg=3$, $p<0.01$). Aus Tabelle 2-2 geht auch hervor, daß in den reduzierten Situationen das faktorielle Design am häufigsten gebraucht wurde, während in den natürlichen Situationen das Kontrollgruppendesign die dominante Versuchsanordnung ist. Die Kriterien für die Klassifikation der Versuchsanordnungen werden im folgenden kurz dargestellt.

(1)Kontrollgruppendesign: Bei dieser Versuchsanordnung werden die Vpn zunächst entweder durch Parallelisierung oder durch Randomisierung gruppiert, dann bleiben die Vpn in der Kontrollgruppe unbehandelt oder erhalten eine andere Behandlung.

Beispiel:



wobei O=Test, P=Parallelisierung, R=Randomisierung,
 VG=Versuchsgruppe, KG=Kontrollgruppe, X_1 =Treatment 1, X_2 =Treatment 2, und X_0 =one Treatment ist.

- (2) Faktorielles Design: Die allgemeine Varianzanalyse, die Kovarianzanalyse und die Varianzanalyse für wiederholte Messungen gehören zu dieser Kategorie. Mit diesen Versuchsanordnungen kann man nicht nur die Haupteffekte der unabhängigen Variablen, sondern auch die Wechselwirkungen zwischen ihnen untersuchen.
- (3) Einzelfall-Experimentaldesign: z.B. A-B-BC-B und A-B-A.
- (4) Randomisiertes Ausbalancierungsdesign: Bei diesem Design werden die Vpn zunächst durch das Verfahren der Randomisierung gruppiert, dann erhalten die Gruppen verschiedene Behandlungen zu verschiedenen Zeitpunkten. Die Reihenfolge der Behandlungen ist für jede Gruppe ausbalanciert. Dieser Kategorie wird auch das Latein-Quadrat-Design zugeordnet.

Beispiel:

$$R \begin{array}{ccc} \nearrow VG_1 & X_1 O & X_2 O \\ \searrow VG_2 & X_2 O & X_1 O \end{array}$$

- (5) Quasi-Experiment: Bei diesem Design gibt es keinen Randomisierungsprozeß bei der Gruppenbildung.

Beispiel:

$$\begin{cases} VG & O & X_1 & O \\ KG & O & X_2 & O \end{cases} ;$$

Solomons Versuchsplan:

$$\begin{cases} G_1 & O & X_1 & O \\ G_2 & O & & O \\ G_3 & & X_1 & O \\ G_4 & & & O \\ G_5 & O & X_2 & O \\ G_6 & & X_2 & O \end{cases}$$

Um die Gegenstände der erziehungswissenschaftlichen experimentellen Forschung zu erfassen und um die Frage zu beantworten, inwieweit die experimentelle Einzelfalluntersuchung in der erziehungswissenschaftlichen Forschung in der BRD angewendet wurde, werden die unabhängigen und

abhängigen Variablen kategorisiert. Die Auftretenshäufigkeiten der Variablen werden im Zusammenhang mit Versuchsanordnungen in Tab. 2-3 und 2-4 zusammengefaßt. Die unabhängigen und abhängigen Variablen werden in je 13 Kategorien klassifiziert. Die operationalen Definitionen und die Zuordnung der Variablen zu den Kategorien lauten folgendermaßen:

A. Die unabhängigen Variablen

- (1) Belastungssituation: bezeichnet "die starke Beanspruchung durch körperliche, geistige und/oder affektive Daueranspannung" (Hehlmann, 1974, S. 50), z.B. Streßsituation; Prüfungsangst; Frustration (Mitteilung von angeblich schlechtem Abschneiden); Belastungssituationen in der Klasse: äußerer Druck vs. Freiwilligkeit (in der Situation mit äußerem Druck wurden die Vpn aufgefordert, die Arbeit in jedem Falle so lange fortzuführen, bis der Versuchsleiter den Schluß bekannt gab. Bei Nachlassen des Arbeitsverhaltens wurden sie ermahnt bzw. zum Weitermachen überredet. In der Situation mit Freiwilligkeit wurde den Vpn gesagt, daß sie nur so lange mitzumachen brauchten, wie sie selbst wollten. Jeder, der keine Lust mehr hatte, durfte die Gruppe bzw. das Zimmer verlassen).
- (2) Einstellung: bezeichnet die verfestigten, bewußten oder unbewußten Meinungen, Vorurteile, Werthaltungen gegenüber Personen oder Sachen, z.B. Zuneigung vs. Furcht beim Kind gegenüber seinem Spielpartner; unbewußte Bewertungs- und Wahrnehmungstendenz gegenüber der eigenen Leistung und der des anderen wurde durch Selbst- vs. Fremdkorrektur bei Rechtschreibdiktaten ermittelt.
- (3) Lernhilfe: bezeichnet die Hinweise, die dem Lernenden zum Verstehen und Lösen der Probleme gegeben wurden. Bei der Forschung der sozialen Wahrnehmung z.B. wurde sie unter folgenden Bedingungen untersucht: (1) vor der Beobachtungsphase mit vs. ohne Aufmerksamkeitslenkung auf die Cues, die für die Regelstruktur sozialer Interaktionen relevant waren, und (2) mit vs. ohne Assoziation dieser Cues mit einem Attribut bei der Beobachtungsphase.
- (4) Modell-Lernen: "Beim Lernen am Modell ahmt der Lernende auf Grund der Beobachtung eines Vorbildes Verhaltensweisen desselben bewußt oder unbewußt nach und macht sie sich zu

Tabelle 2-3: Auftretenshäufigkeit der verschiedenen Kategorien von unabhängigen Variablen, die mit verschiedenen Designs untersucht wurden ¹⁾

Variablen	Versuchsanordnungen					Σ	%
	Kontr. ²⁾	fakt.	Einz.	Ausb.	Quasi-		
1. Belastung	2	2	0	1	1	6	6.3
2. Einstellung	0	0	0	2	0	2	2.1
3. Lernhilfe	1	1	0	0	0	2	2.1
4. Modell-Lernen	3	3	0	0	0	6	6.3
5. Persönlichkeitsvariable	3	13	0	0	0	16	16.7
6. Schulform	4	0	0	0	0	4	4.2
7. Unterrichtsform	7	5	0	1	1	14	14.6
8. Testeffekt	0	1	0	0	1	2	2.1
9. Trainingsprogramm	4	0	0	0	0	4	4.2
10. Unterrichtsmethode	11	4	0	0	1	16	16.7
11. Unterrichtsmittel	5	7	0	2	0	14	14.6
12. Unterrichtsstil	0	1	0	0	0	1	1.0
13. Verhaltensmodifikation	5	1	3	0	0	9	9.4
Σ	45	38	3	6	4	96	

1) Die Seitenangabe der ausgewählten Artikel, in denen die unabhängigen Variablen untersucht worden sind, siehe Anhang 2.

2) Kontr. = Kontrollgruppendesign

fakt. = faktorielles Design

Einz. = Einzelfall-Experimentaldesign

ausb. = randomisiertes Ausbalancierungsdesign

Quasi- = Quasi-Experiment

eigen" (Köck & Ott, 1976, S. 282). Die Faktoren, die den Lernerfolg beim Modell-Lernen beeinflussen, wurden sowohl bezüglich des Vorbildes als auch des Beobachtenden untersucht, z.B. die Kompetenz des Vorbildes (kompetent vs. nicht-kompetent); die Reaktionen des Modells nach der Strafe (ungehorsam-Strafe-keine Folge / ungehorsam-Strafe-gehorsam / ungehorsam-Strafe-ungehorsam / ungehorsam-keine Strafe / gehorsam-keine Strafe / ohne Beobachtung); die Konsequenzen auf das Verhalten des Modells (Regelverstöße-Belohnung / Regelverstöße-ohne Konsequenz / Regelverstöße-Strafe); die Konzentrationsgrade des Beobachtenden; die Fähigkeit des Beobachtenden, die Bedürfnisbefriedigung aufzuschieben; mit vs. ohne Vorkenntnis beim Beobachtenden zur Lösung der Probleme in der Modellverhaltenssituation. Auch der Prozeß des Modell-Lernens wurde untersucht, z.B. Filmbeobachtung ohne Ton / mit Ton / mit Ton plus Verstärkung für die Imitation des kreativen Verhalten; Neben-Sitzen (die hochhängstlichen Kinder wurden in dem Klassenzimmer neben die niedrighängstlichen Kinder gesetzt, um das Angstniveau der hochhängstlichen Kinder zu senken).

- (5) Persönlichkeitsvariable: bezeichnet die individuellen Unterschiede, z.B. das Geschlecht; die Sozialschicht; die Zugehörigkeit zu verschiedenen Leistungsgruppen (Gymnasiasten/Fachschüler/Hauptschüler; kompetenter Gesprächstherapeut vs. unerfahrener Psychologiestudent; Schüler mit hohen mittleren/niedrigen IQ-Werten); die Geschwisterkonstellation der Vpn; die reflexiven vs. impulsiven Vpn.
- (6) Schulform: die Effizienz der verschiedenen Schulformen wurde verglichen, z.B. Vorschule vs. Kindergarten; die musikalische Ausbildung: Normalschule (1 Wochenstunde Musik)/Modellschule "Typ A" (d.h. 1 Wochenstunde Musik + Neigungskurs)/Modellschule "Typ B" (d.h. 5 Wochenstunden Musik + Neigungskurs); vorschulische Einschulung vs. in der Familie bleiben.
- (7) Unterrichtsform: kennzeichnet die Sozialform des Unterrichts und die Lehrform (Köck & Ott, 1976, S. 417), z.B. Kleingruppenarbeit/Einzelarbeit; Partnerarbeit/Einzelarbeit/frontalunterrichtliche Darbietung/frontalunterrichtliche Erarbeitung; Schülerexperiment/Demonstrations-

unterricht/informierender Unterricht; Fernsehmathematik vs. konventioneller Mathematikunterricht; die Kleingruppendiskussion zur Förderung der Verständnisleistung bei den leistungsschwachen Schülern.

(8) Testungseffekt: z.B. mit vs. ohne Vortest; Einstellungstest vs. kritischer Test.

(9) Trainingsprogramm: bezeichnet ein systematisches Programm, das auf die Übung bestimmter Fertigkeiten abzielt, abgesehen von den Trainingskursen in der Schule, die der Kategorie "Unterrichtsmethode" zugeordnet wurden, z.B. das Trainingsprogramm zur Motivänderung (realistische Anspruchsniveausetzung, angemessene Ursachenzuschreibung und positivere Selbst-Bekräftigung); Trainingskurs "Gesprächstherapie" für Erzieherinnen; Encountergruppe-Trainingsprogramm (Ziel dieses Programms ist es, daß sich die Teilnehmer durch spontanen Erfahrungsaustausch mit Repräsentanten verschiedener Schichten oder Berufs- und Interessengruppen auseinandersetzen, um sich besser kennenzulernen und Verständnis füreinander zu entwickeln); mentales Training/praktisches Training/ohne Training für die Bewegungsfertigkeit von Seilspringen (mentales Training bedeutet planmäßig wiederholtes, bewußtes Sich-Vorstellen der zu lernenden sensomotorischen Fertigkeit; praktisches Training heißt tatsächliches Üben dieses Bewegungsablaufes).

(10) Unterrichtsmethode: bezeichnet ein geregeltes Verfahren, nach dem die Unterrichtsstoffe in Unterrichtseinheiten gegliedert werden, und eine Verfahrensweise, mit der die unterrichtlichen Lernprozesse beeinflusst werden (vgl. Köck & Ott, 1976, S. 419), z.B. ganzheitlicher vs. synthetischer Lese-Schreib-Angangsunterricht; Alternativen des Schreiblehrgangs im ersten Schuljahr (sofortiger Beginn mit dem Schreiblehrgang im September/malen zunächst gemischte Antiqua ab und beginnen mit dem Schreiblehrgang im Oktober/malen ebenfalls zunächst gemischte Antiqua ab und beginnen dann mit dem Schreiblehrgang im November); Bielefelder Lese-Rechtschreiblehrgang vs. konventioneller Lese-Rechtschreibunterricht; mit vs. ohne Rechtschreibregeltraining bei Rechtschreibunterricht; das Rechtschreibtraining mit Tonband/mit Tonband + Schablone/ mit Abschreiben

/ohne Training; Effizienzvergleich zwischen verschiedenen Methoden zur Förderung der Entwicklung des Zahlkonstanzbegriffs bei fünfjährigen Kindern: Training der Induktion des kognitiven Konflikts/kombiniertes Konflikt- und Reversibilitätstraining/Training der Mobilisierung des Denkens /Training mit "strukturellem Transfer durch Ausnutzung von Analogie"/ohne Training; die Wirkung des Trainingskurses zur Förderung der Rechtschreibfähigkeit der Legastheniker (Übung mit dem Grundwortschatz und den spezifischen Fehlern des Legasthenikers + Tokenverstärkung); die Wirkung der Vorkenntnisse des Fachrechnenunterrichts; Effizienzvergleich zwischen verschiedenen Methoden des Förderungskurses für die Rechtschreibfertigkeit: Diktat + Fehler-Feedback + soziale Verstärkung/Diktat + Fehler-Feedback/konventioneller Förderungskurs in der Schule/ohne Förderungskurs; Vergleich zweier Modelle von Verbalisierung beim Problemlösungsprozeß: Prozeß- vs. Produktmodell (Prozeßmodell bedeutet, daß der Prozeß, wie die Informationsanalyse, Zielfindung und kreative Problemlösung, Handlungsausführung und Lösungsbewertung vorgenommen wurden, verbalisiert wurde. Beim Produktmodell wurden die Endprodukte, was man alles mit dem Problem machen konnte, verbalisiert); Micro-teaching-Training (fünf Trainingssitzungen mit je einem Erstunterricht und einem verbesserten Zweitunterricht).

- (11) Unterrichtsmittel: beinhaltet die Lehr- und Lernmittel; Lehrmittel werden als Hilfsmittel in den Lehrprozeß eingefügt, um das Vorgetragene und Dargestellte zu verdeutlichen und zu erklären, wie z.B. Abbildungen, Filme, Zeitschriften. Die Lernmittel sind die von dem Lernenden zu verwendenden Gegenstände, z.B. Bücher, Lernprogramme, Arbeitshefte (vgl. Köck & Ott, 1976, S. 242-243): optimierte vs. originale Textversion (Optimierung bedeutet, daß die Textversion im Hinblick auf vier Dimensionen: Einfachheit, Gliederung-Ordnung, Kürze-Prägnanz und zusätzliche Stimulanz zur Erleichterung des Verständnis revidiert wurde); syntaktisch strukturierte vs. unstrukturierte Quasi-Sätze; der Lehr-Text-Umfang: bloße Information/Information + Übungsaufgaben ohne Lösungen/Information + Übungsaufgaben + Lösungen; Lehrtext mit vs. ohne Beschreibung der Ziele;

die Strukturierungshilfe des Studientextes: (a₁) mit Zusammenfassung am Anfang des Textes, (a₂) am Ende des Textes, (a₃) ohne Zusammenfassung, (b₁) mit Unterstreichung, (b₂) ohne Unterstreichung, (c₁) mit Marginalien, (c₂) ohne Marginalien; Lernspiel-Kiste vs. konventionelles vorschulisches Heimprogramm; Tafel und Griffel vs. Heft und Bleistift für die Schreibübung der Erstkläßler; der Effekt des Leistungsprüfgerätes; der Effekt der Trainingsleistung zur Worterkennung bei Vorschulkindern; der Effekt des Schwierigkeitsgrads des Lernstoffs: gut-/mittel-/schwer verständlicher Text; Effekt der Verschiedenheit des Lernstoffs: elektrische Klingel/Elektromotor/Dämmerungsschalter im Physikunterricht.

- (12) **Unterrichtsstil:** ist durch das Ausmaß an Lenkung und Kontrolle des Lernenden durch den Lehrer gekennzeichnet, z.B. autoritärer Unterrichtsstil/demokratischer Unterrichtsstil/Laissez-faire-Stil.
- (13) **Verhaltensmodifikation:** bezeichnet die Verfahren, mit denen auf der Grundlage der Verhaltenstheorie unerwünschte Verhaltensweisen gesenkt und erwünschte Verhaltensweisen aufgebaut werden, z.B. Effizienzvergleich zwischen Gesprächstherapie, autogenem Training und Musiktherapie bei Prüfungsangst; Strafforschung: a) der Effekt der Verzögerung: unmittelbare vs. verzögerte Strafe, b) der Effekt der Mediation: keine Benennung/knappe Erinnerung/ausführliche Erinnerung an die Übertretung); der Effekt der zusätzlichen Zuwendung: mit vs. ohne zusätzlichen Ermutigungssatz neben der Zensur der Klassenarbeit; Ignorieren vs. Response-Cost; soziale Verstärkung vs. soziale + primäre Verstärkung; der Effekt von Time-Out + Ignorieren + sozialer Verstärkung.

B. Die abhängigen Variablen

- (1) **Einstellungsänderung:** wurde durch Einstellungsfragebogen ermittelt, z.B. Fragebogen zur Einstellung gegenüber der Klassenarbeit; der soziale Einstellungstest zur Ermittlung der sozialen Reife und Erwünschtheit; Fragebogen zur Aggressions- bzw. Kooperationsbereitschaft gegenüber anderen Gruppenmitgliedern.

Tabelle 2-4: Auftretenshäufigkeit der verschiedenen Kategorien von abhängigen Variablen, die mit verschiedenen Designs untersucht wurden¹⁾

Variablen	Versuchsanordnungen					Σ	%
	Kontr.	fakt.	Einz.	Ausb.	Quasi-		
1.Einstellung	3	1	0	1	0	5	5.1
2.Emotion	8	2	0	0	0	10	10.2
3.Imitation	1	2	0	0	0	3	3.1
4.Intelligenz	7	1	0	1	0	9	9.2
5.Kreativität	3	0	0	0	1	4	4.1
6.Lehrer- verhalten	2	0	0	0	0	2	2.1
7.Motivation	2	0	0	1	0	3	3.1
8.Psychomoto- rik	2	1	0	0	0	3	3.1
9.Regelver- stöße	0	1	0	0	0	1	1.0
10.Schul- leistung	21	13	0	3	2	39	39.8
11.Sozial- verhalten	3	1	3	1	0	8	8.2
12.Spiel- verhalten	2	1	0	0	0	3	3.1
13.Wahrnehmung	4	3	0	1	0	8	8.2
Σ	58	26	3	8	3	98	

1) Die Seitenangabe der ausgewählten Artikel, in denen die abhängigen Variablen untersucht worden sind, siehe Anhang 3.

- (2) Emotionale Reaktion: Sie bezeichnet im allgemeinen die komplexe affektive Komponente menschlichen Lebens. Sie drückt sich aus durch Lust oder Unlust, Anteilnahme oder Ablehnung, Annäherung oder Vermeidung, Liebe oder Haß (vgl. Köck & Ott, 1976, S. 148). Sie wurde durch Selbsteinschätzungsskala, Angstfragebogen oder Fragebogen zur Erfahrung bei der Schularbeit (die Zufriedenheit mit dem Lernergebnis, oder ob die Vp Spaß beim Lesen hatte) ermittelt.
- (3) Imitationsverhalten: z.B. Imitationsrate oder Übertretungshäufigkeit nach dem Modell-Lernen.
- (4) Intelligenzleistung: Sie umfaßt die geistigen Leistungen, abgesehen von Schulleistungen und Kreativität, z.B. IQ-Wert, Wortschatz, den Begriff der Zahlkonstanz, Formwiedergabe, Mengenerfassung und Abstraktionsfertigkeit.
- (5) Kreativität: Sie bezeichnet "die Fähigkeit und Aufgeschlossenheit zur Innovation" (Köck & Ott, 1976, S. 226). Sie wurde ermittelt durch den "kreativen Gestaltungstest" oder andere Tests, wie z.B. "ungewöhnliche Verwendungen" (kreative Angabe der verschiedenen Verwendungsmöglichkeiten einiger Gegenstände).
- (6) Lehrerverhalten: z.B. einfühlsames Verständnis für die Kinder (sensitiv für die Gefühle und Erlebnisse der Kinder); Wertschätzung der Kinder; stimulierende und nicht-direktive Aktivität gegenüber den Kindern, weiterführendes und anregendes Frageverhalten, Häufigkeit des Bekräftigungsverhaltens, die Einfachheit und Strukturierung des Unterrichtsverhaltens.
- (7) Motivationsänderung: hier beschränkt auf die sekundären Motivationen: "Interesse am dargebotenen Stoff, Konkurrenzsituation unter den Schülern, Hoffnung auf Erfolg, Bedürfnis nach Anerkennung und Geltung, Erreichen des Klassenziels u.a." (Köck & Ott, 1976, S. 284). Sie wurde durch Motivationsfragebogen ermittelt, z.B. der Fragebogen zur "Hoffnung auf Erfolg oder Furcht vor Mißerfolg".
- (8) Psychomotorik: z.B. das aneinanderreihende Malen von Schleifen "llllllll" wurde als ein Indikator der internen Belastungsintensität betrachtet; Schreibmotorik (Häufigkeit von Schreibunterbrechung; Schreibdruck, usw.).

- (9) Regelverstöße: Übertretungshäufigkeit wurde als ein Indikator für Strafeffizienz untersucht.
- (10) Schulleistung: Sie bezeichnet "die Höhe und den Intensitätsgrad des Lernerfolges" (Köck & Ott, 1976, S. 359). Sie wurde durch schriftliche und mündliche Prüfung in standardisierter oder informeller Form ermittelt. Messungen mit Verständnis- und/oder Behaltenstests hinsichtlich des Prozentsatzes der richtig gelösten Aufgaben oder Bearbeitungszeit wurden dieser Kategorie zugeordnet, z.B. Rechtschreibleistung; Musikleistung (z.B. Bentley-Test: Tonhöhenunterscheidung, Tongedächtnis, Akkordanalyse, Rhythmusgedächtnis); Leistungsfähigkeit in Erdkunde, Biologie oder Physik, usw.
- (11) Soziales Verhalten: Es umfaßt alle angemessenen und nicht-angemessenen Verhaltensweisen in einer Gruppenaktivität, z.B. angemessenes Verhalten der Schüler im Unterricht; Dazwischenreden und Nicht-auf-dem-Platz-Sitzen (Störverhalten); Gesprächsverhalten mit den anderen (Kommunikationsfähigkeit); Gewinnaufteilungsverhalten.
- (12) Spielverhalten: z.B. das Bauen von Modellen nach Plan; Seilspringen; Erkennen der Spielstrategien (die Entwicklung des Regelverständnisses von strategischen Spielen bei Kindern von 7-9 Jahren).
- (13) Wahrnehmungsfähigkeit: Sie bezeichnet den Vorgang der Informationsaufnahme aufgrund äußerer Reize oder intrapsychischer Reize. In diesem Vorgang wird der objektive Reizgegenstand zum subjektiv erlebten Wahrnehmungsgegenstand (vgl. Köck & Ott, 1976, S. 440-441). Es gibt sensorische und soziale Wahrnehmung:
- Sensorische Wahrnehmung: z.B. Latenzzeit und Fehler bei "Matching Familiar Figures Test".
- Soziale Wahrnehmung: z.B. Erkennen der zwischenmenschlichen Beziehung (Lernrate und Lerntempo des Beobachtungslernens bei sozialen Interaktionen, d.h. Induktion von Verhaltens~~er~~wartungen); Wahrnehmung von eigenen Leistungen oder Problemen.

Zu Tabelle 2-3 und 2-4 sind einige Anmerkungen zu machen:

- 1) Die Häufigkeiten der verschiedenen Kategorien von Ver-

suchsanordnungen stimmen bei den beiden Tabellen nicht überein (siehe die letzten Zeilen beider Tabellen). Das kommt daher, weil jeder Artikel eine unterschiedliche Anzahl von unabhängigen und abhängigen Variablen hat.

- 2) Tabelle 2-3 zeigt, daß Unterrichtsform, Unterrichtsmethode und Unterrichtsmittel die größten Häufigkeiten (insgesamt 45.9%) haben, und in Tabelle 2-4 die abhängige Variable "Schulleistung" die höchste Häufigkeit hat (39.8%). Dies bedeutet, daß die Unterrichtsforschung ein Schwerpunkt der erziehungswissenschaftlichen experimentellen Forschung ist (vgl. Flehsig, 1976).
- 3) Aus den Tabellen 2-3 und 2-4 ist zu ersehen, daß nur drei experimentelle Einzelfalluntersuchungen referiert wurden, und diese drei Untersuchungen beschränken sich auf die Verhaltensmodifikationsforschung. Dies impliziert, daß die experimentelle Einzelfalluntersuchung von den Erziehungswissenschaftlern in der BRD noch nicht als ein Instrument für die Feldforschung akzeptiert worden ist. Warum konnte die experimentelle Einzelfalluntersuchung nicht über die Verhaltensmodifikationsforschung hinaus in die erziehungswissenschaftlichen Feldforschungen mit einbezogen werden? Ist die experimentelle Einzelfalluntersuchung kein valides Forschungsinstrument? Diese Frage wird im folgenden Abschnitt diskutiert und ist Grundlage der Untersuchung II der vorliegenden Arbeit.

3 Die Validität der experimentellen Einzelfalluntersuchung

3.1 Die Theorie der internen und externen Validität

3.1.1 Die Darstellung der Theorie der internen und externen Validität

Die Anwendbarkeit eines experimentellen Designs hängt wesentlich von seiner Validität ab. Campbell & Stanley (1963, deutsche Übersetzung 1973) haben die Validität von Forschungsdesigns unterschieden in interne und externe Validität.

Interne Validität ist die minimale Voraussetzung eines Experiments und die externe Validität betrifft seine Generalisierbarkeit.

Die Angemessenheit der experimentellen Einzelfalluntersuchung für erziehungswissenschaftliche technologische Experimente muß mit den Kriterien der internen und externen Validität beurteilt werden. Gadenne (1976) hat die von Campbell & Stanley (1973) konzipierte Theorie der internen und externen Validität mit logischen Begriffen weiter entwickelt. Die folgende Darstellung der Theorie der internen und externen Validität spiegelt seine Konzeption wider.

"Die Theorie der internen und externen Validität ist auf alle psychologischen und pädagogischen Untersuchungen anwendbar, in denen es darum geht, eine Hypothese über einen gesetzmäßigen Zusammenhang zwischen Variablen zu testen. Als Spezialfall einer solchen Untersuchung betrachten wir nun einmal ein Experiment zur Prüfung einer Hypothese der folgenden Form: Auf das Ereignis A folgt stets das Ereignis B. Wir schreiben für eine derartige Hypothese abgekürzt $h(A,B)$... Die Realisierung von A wird gewöhnlich als Manipulation der unabhängigen Variablen (UV) bezeichnet, das Eintreten von B als Variation der abhängigen Variablen (AV). In den meisten Experimenten ist B kein qualitatives Merkmal. Um eine solche Veränderung festzustellen, benötigt man in der Regel mindestens zwei Beobachtungen, die miteinander verglichen werden. In bezug auf die Durchführung dieser Beobachtungen bzw. Messungen gibt es zwei verschiedene Vorgehensweisen:

- (1) Die AV wird beim Individuum (oder einer Gruppe) vor und nach der Realisierung von A gemessen. Das Eintreten von B wird durch Vergleich der beiden Messungen festgestellt.
- (2) Zwei vergleichbare Individuen (oder Gruppen) werden in der gleichen Situation den verschiedenen Werten der UV, den experimentellen Bedingungen ausgesetzt. Die AV wird bei beiden Individuen (Gruppen) gemessen. B wird durch Vergleich der Messungen festgestellt.

Im ersten Fall erfolgt die Manipulation der UV intraindividuell, im zweiten interindividuell. Im ersten Fall findet eine Meßwiederholung statt, im zweiten nicht. Eine dieser beiden Strategien oder eine Kombination beider liegt allen Versuchsanordnungen zugrunde." (Gadenne, 1976, S. 11-12).

Das Problem der internen Validität kann nun formuliert werden: "Verursachen die experimentellen Einwirkungen (Behandlungen) wirklich einen Unterschied in dieser besonderen experimentellen Situation?" (Campbell & Stanley, 1973, S. 106). "Der Ausdruck 'wirklich' weist darauf hin, daß es nicht nur darum geht, ob Bu durch Beobachtung bzw. Messung festgestellt werden, sondern auch darum, ob Bu durch Au und nicht etwa durch andere Ereignisse — sogenannte Störfaktoren — bewirkt wurde. Die interne Validität hängt davon ab, ob es gelingt, diese Störfaktoren zu kontrollieren." (Gadenne, 1976, S. 12). Gadenne hat das Prinzip der internen Validität so formuliert: "Eine Untersuchung U ist genau dann intern valide, d.h. man darf genau dann von $Au \wedge Bu$ auf $Au \rightarrow Bu$ schließen, wenn in U alle störenden Bedingungen kontrolliert werden." (Gadenne, 1976, S. 14).

wobei Au = unabhängige Variable in dieser Untersuchung

Bu = abhängige Variable in dieser Untersuchung

$\wedge$ = und

$Au \rightarrow Bu$ = Au ist die Ursache von Bu

"Falls man in einer Untersuchung die Aussage $Au \rightarrow Bu$ akzeptiert, so ist damit über die Gültigkeit der allgemeinen Hypothese $h(A, B)$ zunächst noch nichts ausgesagt. $Au \rightarrow Bu$ bezieht sich auf eine spezielle Untersuchungssituation; $h(A, B)$ behauptet dagegen einen allgemeinen Zusammenhang zwischen A und B. Es stellt sich die Frage, ob man den Befund $Au \rightarrow Bu$ von der Untersuchungssituation auf die Klasse

von Situationen verallgemeinern darf, auf die sich $h(A,B)$ bezieht." (Gadenne, 1976, S. 15). Campbell & Stanley formulieren dieses Problem folgendermaßen: "äußere Gültigkeit wirft die Frage der Generalisierbarkeit auf: für welche Populationen, Gegebenheiten, Behandlungs- und Meßvariablen kann dieser Effekt generalisiert werden?" (Campbell & Stanley, 1973, S. 106).

Zur Beantwortung dieser Frage unterscheidet Gadenne zwischen störenden Bedingungen erster und zweiter Art, wobei die ersten mit der internen Validität, die zweiten mit externen Validität zusammenhängen. "Ein Ereignis F ist eine störende Bedingung 1. Art, wenn F das Ereignis B auszulösen und somit einen Effekt des Ereignisses A vorzutäuschen vermag. Ein Ereignis F' ist eine störende Bedingung 2. Art, wenn A nur zusammen mit F' hinreichend für das Auftreten von B ist. Anders ausgedrückt: Weder A noch F' vermögen allein B zu verursachen, beide zusammen sind jedoch hinreichend für B. Im Unterschied zu F' ist F dazu in der Lage, B allein (d.h. ohne A) zu bewirken." (Gadenne, 1976, S. 15-16). Gadenne formuliert das Prinzip der externen Validität folgendermaßen: "U ist genau dann extern valide, d.h. man darf das Ergebnis $Au \mapsto Bu$ genau dann auf bestimmte Populationen generalisieren, wenn $Au \mapsto Bu$ in bezug auf Stichproben von Individuen, Umgebungen und Variablen akzeptiert werden konnte, die für die entsprechenden Populationen repräsentativ sind." (Gadenne, 1976, S. 19).

Nach Gadenne kann die Theorie der internen und externen Validität als ein 'Zwei-Stufen-Modell' empirischer Untersuchung bezeichnet werden: "Auf der ersten Stufe der Interpretation der Ergebnisse soll eine Entscheidung über die Wahrheit (hier soll die Wahrheit nur als vorläufige Wahrheit oder als eine Annäherung an die Wahrheit betrachtet werden, vom Verfasser) oder Wahrscheinlichkeit singulärer Kausalbehauptungen getroffen werden. Auf der zweiten Stufe soll der Rahmen einer bestimmten Untersuchung durch das Verfahren der Generalisierung überschritten werden." (Gadenne, 1976, S. 42-43).

Zusammenfassend sagt Gadenne (1976): "Eine Untersuchung U ist nur dann eine strenge Prüfung einer Theorie T, wenn in

U alle potentiellen störenden Bedingungen 1. und 2. Art kontrolliert werden." (S.63-64).

3.1.2 Störfaktoren der internen und externen Validität

Campbell & Stanley (1973) haben 8 Störfaktoren der internen und 4 Störfaktoren der externen Validität identifiziert. Die Störfaktoren der internen Validität sind:

- "(1) Zwischenzeitliches Geschehen (history): Die besonderen Ereignisse, die zwischen der ersten und zweiten Messung zusätzlich zu den experimentellen Variablen auftreten.
- (2) Reifung (maturation): Änderungen der Probanden, die eine Funktion des bloßen Zeitablaufes (nicht an besondere Ereignisse gebunden) sind, wie z.B. älter werden, hungrig werden, müder werden, usw.
- (3) Testen (testing): Die Auswirkung einer Testdurchführung auf die Werte einer zweiten Durchführung.
- (4) Hilfsmittel (instrumentation): Insofern als Änderungen in der Beschaffenheit des Meßinstrumentes oder Änderungen der Beobachter oder Auswerter Veränderungen der erhaltenen Messungen hervorrufen können.
- (5) Statistische Regression (regression), die sich auswirkt, wenn Gruppen auf Grund ihrer Extremwerte ausgewählt werden.
- (6) Verzerrende Auswahl (selection): die auf differenzieren der Auswahl der Probanden für die Vergleichsgruppen beruhen.
- (7) Experimentelle Einbußen oder unterschiedlicher Ausfall von Probanden in den Vergleichsgruppen (mortality).
- (8) Wechselwirkung zwischen Auswahl und Reifung usw. (interaction of selection and maturation etc." (Campbell & Stanley, 1973, S.106).

Die Störfaktoren der externen Validität sind:

- "(9) Der reaktive oder Wechselwirkungseffekt des Testens (interaction of testing and X), bei dem eine Voruntersuchung die Empfänglichkeit oder die Antwortbereitschaft der Probanden auf die experimentelle Variable steigern oder verringern könnte. Die an einer voruntersuchten Stichprobe gewonnenen Resultate sind dann hinsichtlich

der Auswirkungen der experimentellen Variablen nicht mehr repräsentativ für die nicht voruntersuchte Population, aus der die Probanden für das Experiment ausgewählt wurden.

- (10) Die Auswirkungen der Wechselwirkung zwischen verzerrter Stichprobenauswahl und der experimentellen Variablen (interaction of selection and X).
- (11) Die reaktiven Effekte der experimentellen Situation (reactive arrangement), die Generalisierungen hinsichtlich der Wirkung der experimentellen Variablen auf solche Personen unmöglich machen, die ihr unter anderen Bedingungen ausgesetzt sind, als sie in der experimentellen Situation herrschten.
- (12) Beeinträchtigung durch mehrere experimentelle Einwirkungen (multiple-X interference), die auftreten kann, wenn dieselben Probanden verschiedenen Behandlungen ausgesetzt werden, denn die Effekte vorhergegangener Behandlungen können nicht immer beseitigt werden. Dies ist besonders bei den Anordnungen mit nur einer Gruppe problematisch." (Campbell & Stanley, 1973, S. 107).

Bracht & Glass haben die externe Validität in zwei Arten klassifiziert: Populations- und Situationsvalidität.

"1. Populationsvalidität

- (a) Experimentell verfügbare Population vs. Zielpopulation: Um Rückschlüsse von der dem Versuchsleiter zur Verfügung stehenden Personengruppe (verfügbare Population) auf die Gesamtmenge der für ihn interessanten Personen (Zielpopulation) möglich zu machen, bedarf es einer gründlichen Kenntnis der Charakteristika beider Populationen. Die Ereignisse beziehen sich möglicherweise nur auf die spezifische Art von Personen, die man zu Vpn machte und nicht auf eine größere Gruppe.
- (b) Interaktion von Persönlichkeitsvariablen und Treatmenteffekten: Es besteht eine Interaktion zwischen Treatmenteffekten und Persönlichkeitsvariablen, wenn die Überlegenheit eines experimentellen Treatments gegenüber einem anderen bei Anwendung auf Versuchspersonen, die sich auf einer anderen Stufe einer Persönlichkeitsvariablen befinden, sich ins Gegenteil verkehrt und

es dem anderen unterlegen wird (dieser Faktor ist gleich dem 10. Störfaktor bei Campbell & Stanley, 1973).

2. Situationsvalidität (ökologische Validität)

- (c) Explizite Beschreibung der unabhängigen Variablen: Verallgemeinerung und Wiederholung experimenteller Ergebnisse setzen ein vollständiges Wissen um alle Aspekte des Treatments und der Versuchsanordnung voraus.
- (d) Interferenz von mehrfachen Versuchseinwirkungen: Wenn zwei oder mehr Treatments nacheinander bei den gleichen Personen in denselben oder anderen Versuchen angewandt werden, so ist es problematisch und manchmal unmöglich, die Ursachen der Ergebnisse zu ermitteln oder sie auf Versuchsanordnungen zu generalisieren, bei denen nur ein Treatment vorhanden ist (gleich dem 12. Störfaktor bei Campbell & Stanley).
- (e) Hawthorne-Effekt: Das Verhalten einer Vp wird zum Teil durch ihre Vorstellung von dem Experiment und durch ihre angestrebte Reaktion auf die experimentellen Stimuli beeinflusst. Das Bewußtsein, an einem Experiment teilzunehmen, kann ein Verhalten erzeugen, das bei einer Versuchsanordnung, die nicht als Experiment wahrgenommen wird, nicht auftreten würde (entspricht dem 11. Störfaktor bei Campbell & Stanley).
- (f) Neuheits- und Störungseffekte: Die experimentellen Ergebnisse können teilweise durch den Enthusiasmus oder die Ablenkung, die durch die Neuheit des Treatments erzeugt werden, entstehen. Der Effekt, der von einem neuen Programm in einer an Veränderungen gewöhnten Umgebung hervorgerufen wird, könnte sehr unterschiedlich von dem in einer veränderungsarmen Umgebung erzeugten Effekt sein.
- (g) Versuchsleitereffekt: Das Verhalten der Vpn kann durch bestimmte Charakteristika oder Verhaltensweisen des Versuchsleiters beeinflusst werden. Die Erwartungen des Versuchsleiters können die Durchführung des Experiments und die Beobachtung des Verhaltens der Vpn beeinflussen.
- (h) Vortest-Sensitivierung: Wenn ein Vortest durchgeführt wurde, können die Ergebnisse des Haupttests zum Teil

eine Folge der Sensitivierung für den Inhalt des Treatments sein. Die Ergebnisse werden u.U. nicht auf eine andere Personengruppe zutreffen, die nicht vorgetestet wurde (gleich dem 9. Störfaktor bei Campbell & Stanley).

- (i) Nachtest-Sensitivierung: Treatmenteffekte können latent oder unvollständig sein und nur dann in Erscheinung treten, wenn ein Nachtest durchgeführt wird.
 - (j) Interaktion zwischen Umweltgeschehen und Treatmenteffekten: Die Resultate können einzigartig sein wegen nicht dazugehöriger Geschehnisse, die während des Experiments vorkommen.
 - (k) Messung der abhängigen Variablen: Die Verallgemeinerung der Ergebnisse hängt von der Bestimmung der abhängigen Variablen und der Wahl der Meßinstrumentarien für diese Variablen ab.
 - (l) Interaktion zwischen dem Zeitpunkt der Messung und den Treatmenteffekten: Messungen der abhängigen Variablen können zu verschiedenen Zeiten verschiedene Ergebnisse erbringen. Ein Treatmenteffekt, den man sofort nach der Durchführung des Treatments beobachten kann, mag einige Zeit später nicht mehr auftreten und vice versa."
- (Bracht & Glass, 1975, S.65-66, die Anmerkungen in den Klammern sind vom Verfasser).

Hersen & Barlow haben die Generalisierung (externe Validität) in der Forschung der Verhaltensveränderung in drei Bereiche gegliedert: die Generalisierung auf Versuchspersonen, Experimentatoren (behavior change agents) und Situationen (vgl. Hersen & Barlow, 1976, S.31-32). Diese grobe Klassifikation entspricht der Klassifikation von Bracht & Glass (1975) mit der Ausnahme, daß Hersen & Barlow den Faktor "Versuchsleitereffekt" aus dem Bereich "Situationsvalidität" ausgliedern.

Kirchner et al. bezweifeln die Brauchbarkeit der Konzeption von Campbell & Stanley für die Psychotherapieforschung und nehmen deshalb eine neue Klassifikation der internen und externen Validität in fünf Bereiche vor: 1. externe Einflüsse auf den Patienten, 2. Patientenmerkmale, 3. Therapeutenmerkmale, 4. Variablen der therapeutischen Technik, und 5. Kriteriumsmessung. (vgl. Kirchner et al. 1977).

Weil ihre Klassifikation von Störfaktoren der internen und externen Validität besonders für die psychotherapeutische Forschung relevant ist, während sich die Arbeit von Campbell und Stanley (1973) speziell auf den Bereich der Unterrichtsforschung bezieht, wird ihre Kritik an dem Ansatz von Campbell & Stanley und ihre neue Konzeption hier nicht weiter berücksichtigt. In der erziehungswissenschaftlichen Forschung ist die von Campbell & Stanley (1973) aufgestellte Störfaktorenliste der internen Validität ein nützliches Kriterium zur Beurteilung der Anwendbarkeit eines Forschungsdesigns. Nur wenn die Störfaktoren durch das Forschungsdesign ausgeschaltet werden, können zuverlässige Aussagen gemacht werden. Aber durch eine einzelne Untersuchung kann die externe Validität nicht abgesichert werden. Wenn man wissen will, welcher Störfaktor zusammen mit dem Treatment die Ergebnisse beeinflußt hat, muß man die gleiche Untersuchung mit anderen Versuchspersonen in anderen Situationen und mit anderen Experimentatoren systematisch wiederholen. Wenn die Ergebnisse einander widersprechen, dann ist die von Bracht & Glass (1975) erwähnte Liste der Störfaktoren der externen Validität nützlich, um den einwirkenden Störfaktor zu identifizieren und Hypothesen für weitere Untersuchungen zu formulieren.

3.2 Die Darstellung der experimentellen Einzelfalluntersuchung

3.2.1 Die Logik der experimentellen Einzelfalluntersuchung

Die grundlegende Struktur der experimentellen Einzelfalluntersuchung ist ABA, wobei A für die Baseline und B für die Behandlung steht. Das Einzelfallexperiment ist besonders nützlich, wenn von vornherein damit gerechnet werden kann, daß der Effekt der experimentellen Variable von vorübergehender oder umkehrbarer Art ist.

Angenommen, eine Wirkung a tritt ein, nachdem eine Behandlung A eingeführt wurde, dann fragt man sich, ob a tatsächlich von A verursacht wird. Empirisch kann die Untersuchung grundsätzlich in drei Phasen aufgeteilt werden:

1. Phase: auf G folgt a'

2. Phase: A wird eingeführt, auf AG folgt a

3. Phase: A wird zurückgeführt, auf G allein folgt wieder a' wobei G = die Gesamtheit der im Experiment bestehenden Sachverhalte (ausgenommen die Ereignisse A, a und a').

A = die vom Versuchsleiter manipulierte unabhängige Variable

a und a' sind die Beobachtungen der abhängigen Variablen und $a \neq a'$

Aus diesem Befund kann man schließen, daß A entweder allein oder zusammen mit G (oder einem Teil von G) die Ursache von a sein muß. A ist für das Eintreten von a notwendig. Die Hypothese, daß G (oder ein Teil von G) allein die Ursache von a sei, kann aufgrund der empirischen Beobachtung ($a \neq a'$) abgelehnt werden. Die Frage, ob A für a nicht nur notwendig sondern auch hinreichend ist, muß durch systematische Variation der Bedingung G beantwortet werden.

Die Logik des Multiplen-Baseline-Designs ist dem ABA-Design ähnlich. Sie lautet wie folgt:

Es gibt beispielsweise drei untersuchte Verhaltensweisen (V_1 , V_2 und V_3) einer Versuchsperson, die voneinander unabhängig sind. Wenn die Veränderung a einer bestimmten Verhaltensweise (V_1 , V_2 oder V_3) nur dann vorkommt, wenn die Behandlung A für jene Verhaltensweise eingeführt wurde und wenn sie sonst immer unverändert bleibt, dann kann man daraus schließen, daß es eine funktionelle Beziehung zwischen der Behandlung und der Verhaltensweise gibt. D.h., A ist für die Veränderung einer Verhaltensweise notwendig, weil die Veränderung einer Verhaltensweise nur nach der Einführung der Behandlung auftritt. Diese funktionelle Beziehung zwischen der Behandlung und der Verhaltensweise kann durch vier Phasen empirischer Untersuchung bewiesen werden:

1. Phase $\begin{cases} V_1: \text{auf G folgt } a_1' \\ V_2: \text{auf G folgt } a_2' \\ V_3: \text{auf G folgt } a_3' \end{cases}$

2. Phase $\begin{cases} V_1: \text{A wird eingeführt, auf AG folgt } a_1 \\ V_2: \text{auf G folgt } a_2' \\ V_3: \text{auf G folgt } a_3' \end{cases}$

3. Phase $\begin{cases} V_1: \text{ auf AG folgt } a_1 \\ V_2: \text{ A wird eingeführt, auf AG folgt } a_2 \\ V_3: \text{ auf G folgt } a_3' \end{cases}$
4. Phase $\begin{cases} V_1: \text{ auf AG folgt } a_1 \\ V_2: \text{ auf AG folgt } a_2 \\ V_3: \text{ A wird eingeführt, auf AG folgt } a_3 \end{cases}$

wobei V_1 = Verhaltensweise 1

V_2 = Verhaltensweise 2

V_3 = Verhaltensweise 3

G, A, a, a' sind genauso definiert wie im oben genannten ABA-Design, und $a_1 \neq a_1'$, $a_2 \neq a_2'$, $a_3 \neq a_3'$.

Wie im Fall des ABA-Designs kann man aus dieser empirischen Untersuchung schließen, daß A entweder allein oder zusammen mit G (oder einem Teil von G) die Ursache von a sein muß. Die Hypothese, daß G (oder ein Teil von G) allein die Ursache von a sei, kann aufgrund der empirischen Beobachtung ($a_1 \neq a_1'$, $a_2 \neq a_2'$ und $a_3 \neq a_3'$) abgelehnt werden. Die Frage, ob A für a nicht nur notwendig sondern auch hinreichend ist, gehört zum Problembereich der externen Validität.

3.2.2 Die Arten (Variationen) der experimentellen Einzelfalluntersuchung

Die Literatur der angewandten Verhaltensforschung ist voll von Beispielen für Ausdehnungen und Variationen des grundlegenden experimentellen ABA-Designs. Die Variationen können in sechs Hauptkategorien klassifiziert werden (Hersen & Barlow, 1976, S. 198ff).

Die erste Kategorie setzt sich aus den Designs zusammen, in denen das AB-Modell wiederholt wird, z.B. ABAB-Design. Dieses Design kann folgendermaßen schematisch dargestellt werden:

$$O_1, O_2, \dots, O_1, X_1 O_{1+1}, X_1 O_{1+2}, \dots, X_1 O_m, O_{m+1}, O_{m+2}, \dots, O_n, \\ X_1 O_{n+1}, X_1 O_{n+2}, \dots, X_1 O_n, \text{ usw.}$$

wobei O = Beobachtung

X = Behandlung

Die Indizes von O bezeichnen die Sitzungen (sessions).

In der zweiten Kategorie werden die verschiedenen Treatmentvariablen mit der Baselinereaktion (baseline performance) im Laufe des Experiments verglichen, z.B. ABACA-Design. Dieses Design kann so dargestellt werden:

$$O_1, O_2, \dots, O_1, X_1 O_{1+1}, X_1 O_{1+2}, \dots, X_1 O_m, O_{m+1}, O_{m+2}, \dots, O_n, \\ X_2 O_{n+1}, X_2 O_{n+2}, \dots, X_2 O_o, O_{o+1}, O_{o+2}, \dots, O_p, \text{ usw.}$$

In der dritten Kategorie werden die spezifischen Variationen der Behandlungsvorgänge während des Experiments überprüft, z.B. A B A B B' B". Dieses Design kann so dargestellt werden:

$$O_1, O_2, \dots, O_1, X_1 O_{1+1}, X_1 O_{1+2}, \dots, X_1 O_m, O_{m+1}, O_{m+2}, \dots, O_n, \\ X_1 O_{n+1}, X_1 O_{n+2}, \dots, X_1 O_o, X_1' O_{o+1}, X_1' O_{o+2}, \dots, X_1' O_p, \\ X_1'' O_{p+1}, X_1'' O_{p+2}, \dots, X_1'' O_q, \text{ usw.}$$

Man blendet z.B. nach der Beweisführung des Kontrolleffekts der Variable im ABAB-Teil dieses Designs den Behandlungsvorgang allmählich aus oder nimmt die Verstärkung zurück. Normalerweise blendet man die Verstärkung sinnvollerweise allmählich aus, um die erreichte Zielverhaltensweise aufrechtzuerhalten.

In der vierten Kategorie werden die Wechselwirkungen oder additiven Effekte von zwei mehr Variablen durch Variationen in dem grundlegenden ABA-Design überprüft, z.B. B-BC-B-A-B-BC-Design. Dieses Design kann so dargestellt werden:

$$X_1 O_1, X_1 O_2, \dots, X_1 O_1, (X_1 + X_2) O_{1+1}, (X_1 + X_2) O_{1+2}, \dots, (X_1 + X_2) O_m, \\ X_1 O_{m+1}, X_1 O_{m+2}, \dots, X_1 O_n, O_{n+1}, O_{n+2}, \dots, O_o, X_1 O_{o+1}, X_1 O_{o+2}, \\ \dots, X_1 O_p, (X_1 + X_2) O_{p+1}, (X_1 + X_2) O_{p+2}, \dots, (X_1 + X_2) O_q, \text{ usw.}$$

Eine solche Analyse wird durch die Prüfung der Effekte der beiden Variablen allein und in Kombination durchgeführt, um

die Wechselwirkung zu bestimmen. Dieses Design kann nicht nur entscheiden, ob die Kombination der beiden Variablen besser ist als die einzelnen Variablen allein, sondern auch analysieren, welche Variable in der Kombination (treatment package) die wesentliche Rolle spielt.

Die fünfte Kategorie ist das Multiple-Baseline-Design, z.B. $\begin{cases} \text{ABBB} \\ \text{AABB} \\ \text{AAAB} \end{cases}$

Dieses Design kann so dargestellt werden:

Verhaltensweise 1: $O_1, O_2, \dots, O_l, X_1 O_{l+1}, X_1 O_{l+2}, \dots, X_1 O_o$

Verhaltensweise 2: $O_1, O_2, \dots, O_m, X_1 O_{m+1}, X_1 O_{m+2}, \dots, X_1 O_p$

Verhaltensweise 3: $O_1, O_2, \dots, O_n, X_1 O_{n+1}, X_1 O_{n+2}, \dots, X_1 O_q$

wobei $l < m < n$

Beim Multiplen-Baseline-Design erfolgt die Behandlung verschiedener Verhaltensweisen jeweils einzeln und unter gleichen Versuchsbedingungen. Ein typisches Beispiel hierfür ist eine Reihe problematischer Verhaltensweisen, die man einzeln behandelt. Der Versuchsleiter versucht nachzuweisen, daß nur das einzelne Verhalten, auf das die Behandlung angewandt wurde, eine Veränderung zeigt; die anderen Verhaltensweisen sollten dagegen unverändert bleiben, bis sie in spezifischer Weise behandelt werden. Dies erfordert natürlich die relative Unabhängigkeit der betreffenden multiplen Verhaltensweisen. Die untersuchten Verhaltensweisen können, wie oben dargestellt, verschiedene Verhaltensweisen einer V_p sein, sie können aber auch die gleiche Verhaltensweise verschiedener V_{pn} oder die gleiche Verhaltensweise in verschiedenen Situationen sein. Dieses Design ist empfehlenswert, wenn von vornherein sicher ist, daß der Effekt der experimentellen Variablen von nicht-vorläufiger oder nicht-umkehrbarer Art ist, eine Situation, in der die Anwendung anderer Einzelfall-Experimentaldesigns problematisch erscheint.

Das sechste Design ist das Multiple-Kontingenz-Design (Multiple schedule design). Dieses Design beruht auf der Annahme, daß ein und dasselbe Verhalten in zwei verschiedenen Situationen einen relativ unabhängigen Verlauf nehmen

kann. Es basiert auf Prinzipien des Diskriminationslernens; d.h. wenn auf dasselbe Verhalten in Gegenwart unterschiedlicher physischer und sozialer Reize verschiedene Konsequenzen folgen, wird es in Abhängigkeit von diesen Reizen unterschiedlich ausfallen, z.B. $\begin{cases} ABAB \\ BABA \end{cases}$

Dieses Design kann so dargestellt werden:

$\begin{cases} \text{Situation 1: } X_1^0, X_1^0, \dots, X_1^0, & X_2^0, X_2^0, \dots, X_2^0, \\ \text{Situation 2: } X_2^0, X_2^0, \dots, X_2^0, & X_1^0, X_1^0, \dots, X_1^0, \\ & \begin{cases} X_1^0, X_1^0, \dots, X_1^0, & X_2^0, X_2^0, \dots, X_2^0 \\ X_2^0, X_2^0, \dots, X_2^0, & X_1^0, X_1^0, \dots, X_1^0 \end{cases} \end{cases}$
wobei X_1 = Reiz 1; X_2 = Reiz 2.

Hersen & Barlow erwähnen noch ein weiteres Design (concurrent schedule design), das der Untersuchungsanordnung 11 (ausbalancierte Anordnung) des Quasi-Experimentaldesigns entspricht. Weil dieses Design sehr selten gebraucht wird (Hersen & Barlow, 1976, S. 226) und es den Störfaktor 8 (Wechselwirkung zwischen Auswahl und Reifung usw.) nicht vollständig ausschalten kann, wird es hier nicht diskutiert.

3.3 Die Anwendbarkeit der von Campbell und Stanley (1973) für die pädagogische Forschung vorgeschlagenen Quasi-Experimente

Campbell & Stanley (1973) haben 10 Quasi-Experimentaldesigns für die pädagogische Feldforschung vorgeschlagen. Darunter besitzen die folgenden vier Anordnungen vollständige interne Validität.

1. Die 8. Anordnung: Die Anordnung mit gleichartigen Zeitstichproben (X_1^0 , X_0^0 , X_1^0 , X_0^0 , usw.)
2. Die 9. Anordnung: Anordnung mit gleichartigem Material ($M_a X_1^0$, $M_b X_0^0$, $M_c X_1^0$, $M_d X_0^0$, usw.)
3. Die 13a. Anordnung: Anordnung mit Stichproben für Vor- und Nachuntersuchung und mit Kontrollgruppen.

{	{	R O (X)
		R X O

	{	R O (X)
		R X O

{	{	R O (X)
		R X O
	{	R O
		R O

	{	R O
		R O

	{	R O
		R O

	{	R O
		R O

4. Die 14. Anordnung: Mehrfache Zeitreihe $\begin{Bmatrix} 0 & 0 & X & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{Bmatrix}$

Die 13a. Anordnung ist bisher noch nicht angewandt worden (Campbell & Stanley, 1973, S. 167), weil sie zahlreiche Gruppen (oder soziale Einheiten) benötigt. Außerdem kann die Gruppengleichheit durch Zufallszuordnung nicht vollständig gewährleistet werden.

Die Anordnungen 8 und 9 sind sehr ähnlich und stehen miteinander in engem Zusammenhang. Der wesentliche Unterschied besteht darin, daß bei Anordnung 9 verschiedene aber gleichartige Materialien für verschiedene Behandlungen und ihre Wiederholungen zusätzlich angewendet werden.

Die Anordnung 8 ist der experimentellen Einzelfalluntersuchung sehr ähnlich. "Die Anordnung (hier die Anordnung 8) kann als Abwandlung des Zeitreihenexperimentes betrachtet werden, bei dem die experimentelle Variable mehrfach wiederkehrt. Das Experiment ist ganz besonders nützlich, wenn von vornherein ziemlich sicher ist, daß der Effekt der experimentellen Variablen von vorübergehender oder umkehrbarer Art ist." (Campbell & Stanley, 1973, S. 153). Die Umkehrung der Behandlung ist auch die grundlegende Eigenschaft des ABAB-Designs. Die Reihenfolge der Behandlungen liegt bei der Anordnung 8 nicht fest, jedoch sollte der zufällige Wechsel dem regelmäßigen vorgezogen werden. "In den meisten Fällen sind einfacher Wechsel der Bedingungen und die Verwendung feststehender Zeiträume nicht wünschenswert; besonders dann nicht, wenn sie sich einem täglichen, wöchentlichen oder monatlichen Zyklus überlagern können oder wenn durch

die vorhersagbare Periodizität eine unerwünschte Konditionierung auf den zeitlichen Zwischenraum stattfindet, die den Unterschied zwischen den beiden Durchführungen verstärken kann." (Campbell & Stanley, 1973, S. 153). Wenn die Reihenfolge der Bedingungen der Anordnung 8 regelmäßig ist, dann entspricht sie dem ABAB-Design mit nur einer Beobachtung in jeder Phase. Wenn die Reihenfolge der Behandlungen randomisiert ist, dann entspricht die Anordnung 8 dem Edgingtonschen Randomisierungstest (Kazdin, 1976, S. 290-298).

Die Anwendbarkeit und Schwäche des Randomisierungstests ist auch von Kazdin (1976) erörtert worden (S. 294-298). Die Anordnung 14 ist von Campbell und Stanley (1973) als eine ausgezeichnete Quasi-experimentelle Anordnung, vielleicht die beste der brauchbaren Anordnungen betrachtet worden. "Da in Schulen wiederholte Messungen leicht vorgenommen werden können, ist die mehrfache Zeitreihenanordnung besonders für die Unterrichtsforschung geeignet." (Campbell & Stanley, 1973, S. 169). Anordnung 14 stellt eine Verbesserung der Anordnung 7 (O O O X O O O) dar. Eine Kontrollzeitreihe wird einbezogen um die Störfaktoren (das Zwischenzeitliche Geschehen und das Hilfsmittel) zu kontrollieren. Der logische Aufbau der Anordnung 14 ist dem Multiplen-Baseline-Design über verschiedene Versuchspersonen (multiple baseline across subjects) (Hersen & Barlow, 1976, S. 239-247) ähnlich, weil der experimentelle Effekt bei der Anordnung 14 auch, wie beim Multiplen-Baseline-Design, zweimal nachgewiesen wird; zum einen durch den Vergleich der verschiedenen Zeitreihen und zum anderen durch den Vergleich der Beobachtungen vor und nach der Durchführung von Intervention in der eigentlichen Reihe (vgl. Campbell & Stanley, 1973, S. 167). Die wesentlichen Unterschiede zwischen den beiden Designs bestehen darin, daß bei der Anordnung 14 das Treatment nur kurze Zeit dauert und keine Beobachtungen während des Treatments durchgeführt werden.

Die Anordnungen 8, 9 und 14, die die interne Validität gefährdenden Störfaktoren beseitigen können, erfordern keine Randomisierung zur Gruppenneubildung und keine große Anzahl von Vpn für eine Untersuchung. Sie sind für die pädagogische Forschung empfehlenswert, weil die interne Validität eine

wichtige Voraussetzung für die Interpretierbarkeit der Ergebnisse wissenschaftlicher Untersuchungen ist.

Diese drei Anordnungen haben aber eine gemeinsame Schwäche: Es können keine Beobachtungen der abhängigen Variable während der Behandlung vorgenommen werden, d.h., der Prozeß der Treatmenteinwirkung wird nicht beobachtet. Dies hat folgende Nachteile: bei der 8. und 9. Anordnung wird in jeder Behandlungsphase nur eine Beobachtung durchgeführt. Falls die Behandlungsphase nur kurz ist, angenommen ein Tag, wird das Auftreten und die Auslöschung der Behandlungseffekte normalerweise nicht so schnell und unmittelbar erfolgen, daß man die Behandlungen jeden Tag wechseln darf (Kazdin, 1976, S. 295). Falls eine Behandlungsphase länger dauert, angenommen eine Woche, und am Ende der Woche die Daten der abhängigen Variable erhoben werden, sind die Daten nur eine einmalige Stichprobe in dieser Woche. Man weiß nicht, wie der Proband innerhalb der Woche reagierte. Bei der Anordnung 14 vergleicht man nur die Unterschiede zwischen Vor- und Nachbehandlung; den Prozeß der Treatmenteinwirkung untersucht man nicht. Da sich die durch eine Behandlung verursachten Änderungen zweifach bemerkbar machen können — zum einen als Niveauänderung (change in level), zum anderen als Trendänderung (change in slope) (vgl. Kazdin, 1976, S. 280-281) — ist es bei der intraindividuellen Untersuchungsanordnung nützlich, den Prozeß der Behandlungseinwirkung zu untersuchen.

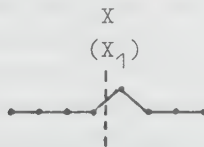
Zusammenfassend läßt sich sagen, daß trotz großer Ähnlichkeiten die experimentelle Einzelfalluntersuchung den Anordnungen 8, 9 und 14 der Quasi-Experimentaldesigns überlegen ist.

3.4 Wie kann die experimentelle Einzelfalluntersuchung die von Campbell & Stanley (1973) erwähnten acht Störfaktoren, die die interne Validität gefährden, ausschalten?

3.4.1 Die Ausschaltung der acht Störfaktoren der internen Validität

Prinzipiell werden die Ursachen für mangelnde interne Validität bei der experimentellen Einzelfalluntersuchung in ähnlicher Weise, im allgemein auch aus denselben Gründen, wie in Anordnung 7 und 8 beim Quasi-Experimentaldesign kontrolliert. Im folgenden wird eingehend zu prüfen sein, wie die von Campbell & Stanley (1973) erwähnten Störfaktoren der internen Validität ausgeschaltet werden können.

- 1) Der erste Störfaktor — zwischenzeitliches Geschehen — hat den stärksten Einfluß auf die interne Validität des Zeitreihendesigns. Nach Campbell & Stanley "ist die mangelnde Kontrolle über das zwischenzeitliche Geschehen die größte Schwäche der Anordnung 7" (Campbell & Stanley, 1973, S.148). Das ABA-Design der experimentellen Einzelfalluntersuchung entspricht im wesentlichen der Anordnung 7, außer, daß man während der Behandlungsphase des ABA-Designs mehrere Beobachtungen durchführt. Kann man durch dieses Vorgehen den Störfaktor " zwischenzeitliches Geschehen" ausschalten? Das zwischenzeitliche Geschehen kann in drei Gruppen klassifiziert werden:
1. Ereignisse, deren Auswirkungen nur von kurzer Dauer sind, wie die Kurve B bei Campbell & Stanley (1973, S.147):



Angenommen, X_1 ist das gleichzeitig auftretende Ereignis, X (Treatment) hat keinen Effekt, X_1 ist für die Veränderung verantwortlich;

2. Ereignisse, die dauerhafte Auswirkung haben, wie die Kurve A bei Campbell & Stanley (1973, S.147);



3. Ereignisse, die zyklische Auswirkung haben, z.B. die Auswirkung der Menstruation der erwachsenen weiblichen Vp auf ihre Emotion; die Auswirkung der Prüfungszeit in der Schule auf das Lernverhalten des Schülers.

Die Auswirkungen der ersten zwei Arten von Ereignissen können beim ABAB-Design identifiziert werden. Die erste Art von Ereignis wird eine plötzliche Schwankung bei der relativ stabilen Kurve in einer Phase verursachen. Die zweite Art von Ereignis wird eine weitergehende Wirkung zeigen, obwohl das Treatment schon nach dem Phasenwechsel zurückgezogen worden ist. Die dritte Art von Ereignis kann nicht leicht identifiziert werden, sie muß vermieden werden. Man soll z.B. die Behandlung nicht in den Zeitraum von Prüfungen legen. Hersen & Barlow haben für solche Fälle vorgeschlagen, die Beobachtungsphasen auszuweiten (vgl. Hersen & Barlow, 1976, S. 105).

Das ABAB-Design hat große Ähnlichkeit mit der Anordnung 8, die bei vorhersagbarer Periodizität eine unerwünschte Konditionierung auf den zeitlichen Zwischenraum ermöglicht. Dieser Effekt kann in Anordnung 8 durch Randomisierung der Reihenfolge von X_1O und X_0O (Campbell & Stanley, 1973, S. 153), und beim ABAB-Design durch die Elastizität der Phasenlänge vermieden werden.

- (2) Wenn Reifung (2. Störfaktor), Übung oder Vortest (3. Störfaktor) in der Untersuchung eine Rolle spielen, wird man feststellen daß in der nächsten Phase keine Umkehrung möglich ist, weil solche Wirkungen normalerweise dauerhaft sind und sich in die nächste Phase erstrecken.
- (3) Der Effekt der Veränderung der Meßgeräte (4. Störfaktor) zeigt sich in einer plötzlichen Schwankung der Beobachtungsdaten. Während des Experiments dürfen die Meßgeräte nicht willkürlich geändert werden. Bei plötzlichen Schwankungen in den Beobachtungsdaten muß man überprüfen, ob diese auf Störungen oder Versagen der Meßgeräte zurückzuführen sind. Wenn in den Meßvorgang Beurteilungen von

Beobachtern eingehen, muß die Zuverlässigkeit der Beobachtungen gewährleistet sein, sonst steht die Validität der Ergebnisse in Frage. Deshalb braucht man bei der Verhaltensbeobachtung mindestens zwei unabhängige Beobachter. Ein Maß für die Zuverlässigkeit ist der Grad der Übereinstimmung zwischen den Beobachtern.

- (4) Regression (5. Störfaktor), verzerrte Auswahl von Vpn (6. Störfaktor), verschiedene Einbuße von Vpn der zu vergleichenden Gruppen (7. Störfaktor) und Wechselwirkung zwischen Auswahl und Reifung usw. (8. Störfaktor) spielen nur bei den interindividuellen Designs (beim Gruppenvergleich) eine Rolle. Für Einzelfallexperimente haben sie keine Bedeutung. Nur bei N=1-Gruppe-Design "sollte ein Bericht über Abwesenheiten, vorzeitiges Ausscheiden und Ersatz von Probanden aufgenommen werden, damit sichergestellt ist, daß die zufälligen personalen Veränderungen keine einleuchtenden Alternativhypothesen bieten." (Campbell & Stanley, 1973, S.150).

3.4.2 Ein weiterer Störfaktor, der die interne Validität möglicherweise gefährden kann

Ein anderer potentieller Störfaktor, der die interne Validität gefährden kann und nicht von Campbell & Stanley erwähnt wird, ist die Schwankung der Beobachtungen. Bei den Gruppendesigns werden die Unterschiede zwischen den Experiment- und Kontrollgruppen statistisch überprüft. Dadurch wird die alternative Interpretation — die Unterschiede könnten auf die zufälligen Datenschwankungen zurückgeführt werden — irrelevant. Bei der experimentellen Einzelfalluntersuchung erfolgt die Beurteilung der Unterschiede zwischen den Behandlungs- und Baselinephasen intuitiv. In den meisten Forschungsberichten im Journal of Applied Behavior Analysis vergleicht man nur die Mittelwerte der verschiedenen Phasen ohne Anwendung statistischer Prüfverfahren (darauf wird in Untersuchung II eingegangen). Wenn die Kurven nicht stabil sind, d.h. wenn die Schwankung in den Beobachtungen groß ist, wird die interne Validität der Untersuchung fragwürdig. Dieser Faktor ist ein Störfaktor der internen Validität, weil sich durch ihn allein, d.h.

ohne Annahme eines Effekts der unabhängigen Variable, die Ergebnisse interpretieren lassen.

Campbell & Stanley (1973) sagen: " In den fortgeschrittenen Wissenschaften werden Signifikanztests seltener benutzt als in der Psychologie und Pädagogik, weil die Stärke und Eindeutigkeit der Effekte, mit denen sie sich beschäftigen, Signifikanzprüfungen überflüssig machen. Wenn die konventionellen Signifikanztests benutzt würden, ergäben sich hochsignifikante Werte. Es ist jedoch für die Ökologie der Sozialwissenschaften typisch, daß mit viel weniger eindeutigen Werten gearbeitet werden muß, die eine Signifikanzprüfung erfordern. Immer wenn gesunder Menschenverstand und intuitive Einsicht einen eindeutigen Effekt erwarten lassen, ist wahrscheinlich auch ein Signifikanztest durchführbar, der die dem intuitiven Urteil zugrundeliegenden Betrachtungen objektiviert." (S.151-152). Sind die Effekte der Behandlungen bei der experimentellen Einzelfalluntersuchung so stark und eindeutig genug, daß sie die statistische Prüfung erübrigt? Dieses Problem ist ein Gegenstand der Untersuchung II der vorliegenden Arbeit.

3.5 Wie kann die experimentelle Einzelfalluntersuchung externe Validität durch systematische Wiederholung erreichen?

3.5.1 Der Zusammenhang zwischen systematischer Wiederholung und externer Validität

Man kann in einem Experiment zeigen, daß A für das Zustandekommen von a notwendig, nicht aber, daß es hinreichend ist. Auch wenn A beim Auftreten von a immer anwesend und beim Ausbleiben von a immer abwesend sein sollte, wäre es doch denkbar, daß A nur zusammen mit weiteren Faktoren a hervorzubringen vermag. Ob es solche Faktoren gibt oder nicht, muß durch systematische Variation der Bedingungen nachgewiesen werden. Dies ist jedoch genau die Strategie, die dem Prinzip der externen Validität zugrunde liegt: Ein spezieller Kausalzusammenhang zwischen einer unabhängigen Variablen und einer abhängigen Variablen wird unter Variation weiterer Bedingungen untersucht. Sollte der Zusammenhang unter allen

Bedingungen nachweisbar sein, so darf er generalisiert werden. Sollte der Zusammenhang dagegen unter bestimmten Bedingungen nicht nachweisbar sein, so ist eine Generalisierung nicht erlaubt. Aber wenn der Grund der Nicht-Nachweisbarkeit dieses Zusammenhangs unter bestimmten Bedingungen empirisch interpretiert werden kann, d.h., wenn der Störfaktor, der mit der Behandlung zusammen den Effekt bewirkte, empirisch analysiert und bestimmt werden kann, dann wird die Generalisierung nicht abgelehnt, sondern stößt auf ihre Grenze.

Snow (1974) meint, daß die systematische Wiederholung der andere Schritt zur Repräsentativität ist. Ein Forscher kann nicht nur die Vpn variieren, um die Populationsrepräsentativität zu untersuchen (Sidman (1960) spricht in diesem Zusammenhang von direkter Wiederholung), sondern darüberhinaus auch die Lehrer, die Methoden, Kontexte oder die Lernmaterialien variieren, um die ökologische Repräsentativität zu untersuchen.

3.5.2 Die direkte Wiederholung

Wiederholung ist der Kern jeder Wissenschaft. In allen Wissenschaften dient die Wiederholung mindestens zwei Zielen: erstens, die Reliabilität der vorhergehenden Befunde zu bestätigen, zweitens, die Generalisierung dieser Befunde unter verschiedenen Bedingungen festzustellen.

Für die experimentellen Einzelfalluntersuchung schlägt Sidman (1960) vor, daß, bevor man die systematische Wiederholung vornimmt, man zunächst die direkte Wiederholung durchführt. Bei der direkten Wiederholung wird eine Behandlung durch denselben Versuchsleiter oder dieselbe Gruppe von Versuchsleitern in einer spezifischen Situation mit einer Reihe von Vpn, die bezüglich eines besonderen Merkmals homogen sind, durchgeführt.

Hersen & Barlow haben folgende Richtlinien für die Durchführung der direkten Wiederholung vorgeschlagen:

- (1) Während der Wiederholungen sollen die Versuchsleiter und die Situationen konstant bleiben.
- (2) Die untersuchten, abhängigen Variablen bei verschiedenen Vpn sollen topographische Ähnlichkeit haben.
- (3) Die Störvariablen (background variables) der Vpn (z.B.

Alter, Geschlecht, usw.) sollen ähnlich, wenn nicht identisch sein.

- (4) Der Behandlungsvorgang soll für jede Vp solange gleich sein, bis das Experiment fehlschlägt. Wenn das Experiment nicht erfolgreich verläuft, soll man versuchen, durch improvisierte und sich schnell ändernde experimentelle Designs die Ursache der interindividuellen Variabilität festzustellen (vgl. Hersen & Barlow, 1976, chap. 2 section 2.3). Wenn es dem Forscher gelingt, die Quelle der Variabilität zu identifizieren, soll er diese neuen Ergebnisse mit anderen Vpn, die die gleichen Charakteristika haben, überprüfen. Wenn es ihm nicht gelingt, muß er die Unterschiede zwischen dieser Vp und den anderen Vpn, mit denen er erfolgreiche Untersuchungen gemacht hat, feststellen, um weitere Untersuchungen durchführen zu können.
- (5) Ein erfolgreiches Experiment und drei erfolgreiche direkte Wiederholungen sind in der Regel ausreichend, um die systematische Wiederholung der topographisch unterschiedlichen Verhaltensweisen in derselben Situation oder dasselbe Verhalten in verschiedenen Situationen zu beginnen. Wenn dagegen die direkte Wiederholung keine eindeutigen Ergebnisse zeigt, muß er darüber entscheiden, wann er eine neue Untersuchungsreihe beginnt, um die Quelle der Variabilität festzustellen. (vgl. Hersen & Barlow, 1976, S. 334-335).

3.5.3 Die systematische Wiederholung

Die systematische Wiederholung bedeutet, daß die Befunde aus einer Reihe von direkten Wiederholungen bei verschiedenen Situationen, Versuchsleitern und abhängigen Variablen oder jeder Kombination davon wiederholt werden. Es wird deutlich, daß jede erfolgreiche systematische Wiederholungsreihe, in der ein oder mehrere der obengenannten Faktoren variiert wurden, weitere Information über die Populationsvalidität ergibt, weil bei solchen Prozessen gewöhnlich neue Vpn einbezogen werden.

Hersen & Barlow haben folgende Richtlinien für die Durchführung der systematischen Wiederholung vorgeschlagen:

- (1) Der Forscher, der sich für die systematische Wiederholung entschieden hat, soll die Unterschiede zwischen seinem und dem originalen Experiment bezüglich der Vpn, Versuchsleiter und Situationen so deutlich hervorheben, daß andere Forscher zusammen mit ihm eine Hypothese über die Abweichungen aufstellen können, falls die Untersuchung keinen Erfolg zeigt. In einer systematischen Wiederholung werden normalerweise nur eine oder möglicherweise zwei Variablen variiert. Aber der Forscher kann auch mehr Störvariablen in einer Untersuchung variieren. Wenn das Experiment erfolgreich ist, hat er dadurch einen großen Vorsprung, wenn das Experiment aber fehlschlägt, kann er nicht feststellen, welche veränderte Variable oder Kombination von Variablen für das Scheitern verantwortlich ist, und er muß die Versuchsreihe wiederholen.
- (2) Die systematische Wiederholung ist im wesentlichen eine Suche nach Ausnahmefällen (a search for exceptions). Wenn keine Ausnahme während der Wiederholungen auftritt, wird die allgemeine Gültigkeit der Theorie bestätigt. Das Ziel der systematischen Wiederholung ist aber, die Grenzen der Theorie aufzuzeigen, d.h., die Bedingungen festzustellen, in denen die Versuche Erfolg haben oder nicht.
- (3) Es stellt sich schließlich die Frage, wann die systematische Wiederholungsreihe zu beenden ist. In der angewandten Forschung ist die systematische Wiederholungsreihe, wie die in der Grundlagenforschung, unendlich, weil die Wissenschaftler immer die Ausnahmefälle von einem gegebenen Prinzip finden sollen. (vgl. Hersen & Barlow, 1976, S. 353-355).

In der Tat gibt es wenige Wissenschaftler, die die Experimente absichtlich um der systematischen Wiederholung willen durchführen (vgl. Sidman, 1960, S. 137). Obwohl die systematische Wiederholung unschätzbar und notwendig für wissenschaftlichen Fortschritt ist, tritt sie zum großen Teil nur als Nebenprodukt des Interesses des Forschers in seinem Forschungsbereich auf. Der Forscher führt die Experimente durch, um die natürlichen Phänomene zu analysieren und ihre gegenseitigen Beziehungen festzustellen. Nachdem er in

einem Experiment die Kontrolle über die unabhängigen und abhängigen Variablen erreicht hat, fängt er selbst an, oder veranlaßt er durch seine Veröffentlichung andere Forscher, weitere Variablen in diesem Bereich zu erforschen. Unter diesem Umstand ist die systematische Wiederholung, ohne bewußte Berücksichtigung von Seiten der Forscher, eine unvermeidliche Begleiterscheinung der systematischen Forschung. Wenn die Befunde der ursprünglichen Experimente bestätigt werden, nehmen ihre Reliabilität und Generalisierung zu, wenn Diskrepanzen vorkommen, lautet die überwältigende Schlußfolgerung, daß die originalen Experimente einige einflußreichen Variablen nicht beachtet haben. Aber die Ergebnisse der originalen Experimenten werden nicht als falsch, sondern als spezifische Fälle innerhalb eines neuen allgemeineren Rahmens betrachtet (vgl. Sidman, 1960, S.133-134).

4. Untersuchung II: Die Probleme in der Durchführung der experimentellen Einzelfalluntersuchung

(a) Probleme bezüglich der inneren Validität:

Die wesentlichen Bestandteile der experimentellen Einzelfalluntersuchung sind die Stabilität und die Umkehrbarkeit (reversibility) der Baseline. Bei diesem Design gibt es eine anfängliche Baseline, in der ermittelt wird, wie häufig das spezifische Verhalten unbehandelt vorkommt. Diese Basis-messungen werden im allgemeinen so lange fortgesetzt, bis sich ein stabiles Muster ergibt. Die Baseline wird als A-Phase bezeichnet. Im Anschluß an die Baseline wird eine therapeutische Variable oder Behandlung eingeführt. Die Einführung der Behandlungsvariablen stellt die B-Phase des Experiments dar. Auch diese Phase wird solange fortgesetzt, bis sich in den Meßdaten ein stabiles Muster bzw. eine stabile Tendenz zeigt. Dann setzt man die Behandlung ab, um in einer zweiten A-Phase durch die Rückkehr zu den Ausgangsbedingungen eine Annäherung an die Baseline zu erreichen. Diese Umkehrung in der zweiten A-Phase soll sicher stellen, daß die Veränderungen im Zielverhalten aus der Einführung bzw. Absetzung der Behandlungsvariablen resultieren.

Daraus ergeben sich zwei Fragen:

- (1) Gibt es ein objektives Kriterium für die Stabilität?
- (2) Wie kann man die Verhaltensänderung in den beiden benachbarten Phasen objektiv beurteilen?

Für die Stabilität haben Schoenfeld et al. (1956) folgendes Kriterium vorgeschlagen:

$$\frac{\sum_{i=n-5}^{n-3} X_i/3 - \sum_{i=n-2}^n X_i/3}{\sum_{i=n-5}^n X_i/6} < 0.05 \text{ (absoluter Wert)}$$

wobei X_i = Beobachtungen

n = Die Anzahl der Beobachtungen in einer Phase

Wenn diese Kriterium nicht erreicht wird, führt man noch eine Beobachtung durch und überprüft mit dem folgenden

Kriterium:

$$\frac{\sum_{i=n-6}^{n-1} X_i / 6 - X_n}{\sum_{i=n-6}^n X_i / 7} < 0.05 \text{ (absoluter Wert)}$$

Es setzt mindestens sechs Beobachtungen in jeder Phase voraus. Dieses Kriterium ist bisher nicht akzeptiert und verwendet worden, weil bei diesem Kriterium nur die Mittelwerte der Beobachtungen berücksichtigt werden. Die Variabilität der Beobachtungen wird vernachlässigt. Bei der folgenden hypothetischen Kurve ist z.B. der Mittelwert der ersten drei Beobachtungen gleich dem Mittelwert der letzten drei und die Differenz zwischen diesen beiden Mittelwerten beträgt weniger als fünf Prozent des Mittelwerts der sechs Beobachtungen; trotzdem kann man nur schwer akzeptieren, daß diese Kurve stabil ist.

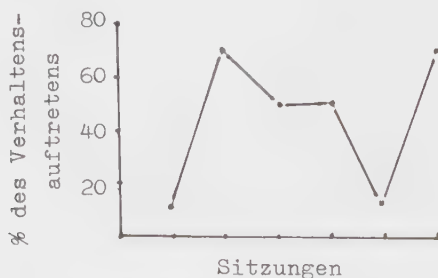


Abb. 4-1: Eine hypothetische nicht-stabile Kurve

Bei den meisten Forschungsberichten in den untersuchten Zeitschriften werden die Stabilität und Veränderung des Verhaltens nur durch visuelle Beurteilung überprüft. Die visuelle Überprüfung ist aber nicht objektiv.

In einer Untersuchung über die Übereinstimmung zwischen den visuellen Beurteilungen und den Zeitreihenanalyse für die Überprüfung der Differenz zwischen den benachbarten Phasen haben Jones et al. (1978) gefunden, daß sowohl die Interbeurteilerübereinstimmung als auch die Übereinstimmung

zwischen den visuellen Beurteilungen und den Zeitreihenanalysen niedrig waren, wenn die Zeitreihenabhängigkeit der Beobachtungen hoch war.

Insgesamt 58 Phasenpaare, die keine einschneidende Differenz zeigten, aber von den Autoren als signifikant verschieden angesehen wurden, wurden den 11 Beurteilern präsentiert. Diese 11 Beurteiler (Versuchspersonen) — Vollzeit-Forscher, Universitätsprofessoren oder graduierte Studenten mit 3 bis 17 jähriger Forschungserfahrung in Psychologie, einschließlich der angewandten Verhaltensanalyse — sollten entscheiden, ob es von einer Phase zur anderen eine signifikante Niveau-Änderung gab. Die Ergebnisse zeigen, daß die Interbeurteilerübereinstimmungen für 58 Phasenpaare zwischen 0.04 und 0.79 mit einem Median von 0.39 rangieren. Dies wirft Zweifel an der Zuverlässigkeit der visuellen Beurteilung in dieser Studie auf.

Kazdin behauptet, daß das experimentelle Kriterium erreicht sei, wenn sich die Beobachtungen der Behandlungsphase nicht mit denen der Baseline überschneiden (overlap), und wenn es divergente Slopes in Behandlungs- und Baselinephase gebe, deshalb erübrige sich eine statistische Überprüfung (vgl. Kazdin, 1976, S. 268). Das erste Kriterium kann akzeptiert werden, weil es objektiv ist, aber das zweite bedarf noch statistischer Überprüfung, weil die Tendenz wegen der Variabilität der Beobachtungen manchmal nicht klar ersichtlich ist. Um die Verhaltensänderungen objektiv zu überprüfen, ist zu fordern, daß statistische Prüfverfahren durchgeführt werden, wenn sich die Beobachtungen beider benachbarten Phasen überschneiden.

(b) Probleme hinsichtlich der äußeren Validität:

Man definiert die äußere Validität als die Übertragbarkeit der experimentellen Ergebnisse auf andere Populationen, Situationen (settings), Beobachtungsvariablen und Behandlungsvariablen. Inwieweit hat die experimentelle Einzelfalluntersuchung hinsichtlich der o.g. Variablen Anwendung gefunden?

Solche Probleme werden im folgenden untersucht und diskutiert.

4.1 Fragestellungen

- (a) Fragestellungen hinsichtlich der inneren Gültigkeit:
- (1) Wieviel Phasenpaare in den veröffentlichten Forschungsberichten haben das Kriterium des Nicht-Überschneidens nicht erreicht? Sind diesbezüglich irgendwelche statistischen Prüfverfahren angewandt worden?
 - (2) Wie groß ist die Anzahl der Beobachtungen in jeder Phase? Barlow & Hersen (1973) behaupten, daß in jeder Phase mindestens drei Beobachtungen durchgeführt werden müssen. Ist das genug? Wie haben es die meisten Forscher gemacht?
 - (3) Wie hoch soll die prozentuale Übereinstimmung zwischen den Beobachtern sein? Wie haben es die meisten Forscher gemacht?
- (b) Fragestellungen hinsichtlich der äußeren Gültigkeit:
Inwieweit ist bisher die experimentelle Einzelfalluntersuchung hinsichtlich der Versuchspersonen, Experimentatoren, Situationen, unabhängigen und abhängigen Variablen verallgemeinert worden?

4.2 Methode

4.2.1 Sammlung der Daten

Nach den in Tabelle 4-1 dargestellten Items werden Informationen aus den Forschungsberichten der ausgewählten Zeitschriften im Bereich der Verhaltensmodifikation gesammelt.

Tabelle 4-1: Items für die Sammlung der Informationen

-
- (A) Titel
- (B) Versuchspersonen (Alter, Geschlecht, behindert oder nicht-behindert)
- (C) Situationen (Klinik, Schule, usw.)
- (D) Experimentatoren (Therapeuten, Lehrer, Eltern, usw.)
- (E) Designarten
- (F) Unabhängige Variablen
- (G) Abhängige Variablen
- (H) Beobachtungen in den Phasen
- 1) Die Anzahl der Beobachtungen jeder Phase.

- 2) Wieviel Paare der benachbarten Phasen haben Überschneidungen in den Beobachtungen? Wieviel haben keine Überschneidungen?
 - 3) Ist irgendein statistisches Prüfverfahren angewandt worden?
 - 4) Wie hoch ist die Interbeobachterübereinstimmung?
-

4.2.2 Auswahl der Zeitschriften

Vier Zeitschriften werden für die Untersuchung ausgewählt:

- (1) Journal of Applied Behavior Analysis (JABA) (1968-1976)
- (2) Behavior Therapy (BT) (1970-1976)
- (3) Behavior, Research and Therapy (BRT) (1970-1976)
- (4) Journal of Behavior Therapy and Experimental Psychiatry (JBTEP) (1970-1976)

Die Gründe für diese Auswahl lauten:

- (1) Diese vier Zeitschriften werden von Kazdin (1975) als Zeitschriften für Verhaltensmodifikation klassifiziert. Etwa 90% der in dem Journal of Applied Behavior Analysis veröffentlichten Untersuchungen liegt die Technik der angewandten Verhaltensanalyse (m.a.W. der experimentellen Einzelfalluntersuchung) zugrunde, und die anderen drei Zeitschriften enthalten ca. 30% solcher Forschungsberichte.
- (2) Böttger et al. (1977) verschickten 50 Fragebögen an praktisch-therapeutisch oder wissenschaftlich tätige Verhaltenstherapeuten in der Bundesrepublik mit der Bitte, sechs deutsch- oder fremdsprachige Zeitschriften anzugeben und diese entsprechend ihrer wissenschaftlichen Bedeutung für die Verhaltenstherapie in eine Rangreihe zu bringen. Als Ergebnis der Befragung erhielten diese vier Zeitschriften die höchsten Bewertungen.

4.2.3 Kriterien für die Aufnahme von Forschungsberichten

Da die ausgewählten Zeitschriften für Verhaltensmodifikation eine Reihe sehr verschiedener Beiträge enthalten, werden in der Untersuchung folgende Kriterien für die Auswahl der Artikel festgelegt:

- (1) In dem ausgewählten Artikel muß mindestens eine experi-

mentelle Einzelfalluntersuchung angewandt worden sein.

- (2) Die Untersuchung muß in irgendeiner Form eine Intervention zur Verhaltensmodifikation beinhalten.
- (3) Vor-experimentelle Studien, z.B. die historische Einzelfallstudie (case study) oder das A-B-Einzelfall-Experimentaldesign, werden wegen der Mängel bezüglich der internen Validität nicht aufgenommen, da bei solchen Designs alternative Interpretationen des Treatmenteffekts nicht ausgeschaltet werden können.
- (4) Artikel, die die Verhaltensrate kumulativ (cumulative record) darstellen, werden nicht zur Analyse herangezogen, weil sie keine Information für das Item H der Tabelle 4-1 liefern können.
- (5) Untersuchungen mit Tierexperimenten werden ausgeschlossen.

4.3 Ergebnisse

Die Anzahl der ausgewählten Artikel werden für die einzelnen Zeitschriften in den Tabelle 4-2, 4-3, 4-4 und 4-5 angegeben.

Tabelle 4-2: Prozentsätze der ausgewählten Artikel aus
Journal of Applied Behavior Analysis (JABA)

Jahrgänge	'68	'69	'70	'71	'72	'73	'74	'75	'76	Summe
Gesamtzahl der Artikel ¹⁾	42	37	41	38	65	79	89	61	82	534
Ausgewählte Artikel	21	17	26	18	26	35	42	26	34	245
% ²⁾	50.0	45.9	63.4	47.4	40.0	44.3	47.2	42.6	41.5	45.9

1) Die Gesamtzahl der Artikel jedes Jahrgangs ist aus der Inhaltsgliederung jedes Jahrgangs errechnet worden.

2)
$$\% = \frac{\text{Ausgewählte Artikel}}{\text{Gesamtzahl der Artikel}} \cdot 100$$

Tabelle 4-3: Prozentsätze der ausgewählten Artikel aus
Journal of Behavior Therapy and Experimental
Psychiatry (JBTEP)

Jahrgänge	'70	'71	'72	'73	'74	'75	'76	Summe
Gesamtzahl der Artikel	61	59	74	70	86	108	113	571
Ausgewählte Artikel	3	4	5	8	6	18	11	55
%	4.9	6.8	6.8	11.4	7.0	16.7	9.7	9.6

Tabelle 4-4: Prozentsätze der ausgewählten Artikel aus
Behavior Therapy (BT)

Jahrgänge	'70	'71	'72	'73	'74	'75	'76	Summe
Gesamtzahl der Artikel	91	115	141	150	146	168	153	964
Ausgewählte Artikel	1	5	5	9	7	9	13	49
%	1.1	4.3	3.5	6.0	4.8	5.4	8.5	5.1

Tabelle 4-5: Prozentsätze der ausgewählten Artikel aus
Behavior, Research and Therapy (BRT)

Jahrgänge	'70	'71	'72	'73	'74	'75	'76	Summe
Gesamtzahl der Artikel	63	51	58	52	88	48	75	435
Ausgewählte Artikel	0	1	1	2	7	4	3	18
%	0.0	2.0	1.7	3.8	8.0	8.3	4.0	4.1

Insgesamt wurden 367 Artikel ausgewählt. Wegen der geringen Anzahl der ausgewählten Artikel aus den drei Zeitschriften, JBTEP, BT und BRT, werden deren Daten nicht pro Jahrgang, sondern in bezug auf den gesamten Untersuchungszeitraum 1970-1976 analysiert.

Im folgenden werden die Ergebnisse in der Reihenfolge der oben aufgeworfenen Fragestellungen präsentiert.

(a) Die Fragestellungen hinsichtlich der internen Validität:

- (1) Wieviel Phasenpaare in den veröffentlichten Forschungsberichten haben das Kriterium des Nicht-Überschneidens nicht erreicht? Sind diesbezüglich irgendwelche statistischen Prüfverfahren angewandt worden?

Das Kriterium des Nicht-Überschneidens ist dann erreicht, wenn der niedrigste Datenpunkt einer Phase höher als der höchste Datenpunkt der benachbarten Phase ist. Bei der Berechnung der sich nicht-überschneidenden Phasenpaare werden zwei Arten von Daten ausgeschlossen: 1. Wenn N = eine Gruppe und die Gruppe als eine Einheit behandelt wird, werden nur die Gruppendaten berechnet, die Daten von einzelnen Versuchspersonen werden nicht berücksichtigt. 2. Die Daten von den Follow-up- und Generalisierungsphasen werden nicht berechnet.

Tabelle 4-6 gibt die Prozentsätze der sich nicht-überschneidenden Phasenpaare an. Es zeigt sich, daß nur 32,5% der untersuchten Phasenpaare das Kriterium von Nicht-Überschneiden erreichen, m.a.W., bei einer experimentellen Einzelfalluntersuchung ist nur mit ca. 1/3 Wahrscheinlichkeit das Auftreten von sich nicht-überschneidenden Phasenpaaren zu erwarten. Abb. 4-2 zeigt, daß die Variabilität dieser Wahrscheinlichkeit nicht groß ist (schwanken zwischen 25.6% und 39.7%, $\sigma = 4.32\%$).

Die übrigen 2/3 der Phasenpaare bedürfen objektiver Überprüfung. Um die Interpretation der Forschungsergebnisse der experimentellen Einzelfalluntersuchung zu objektivieren, ist es deshalb empfehlenswert, statistische Prüfverfahren anzuwenden.

Unter den 367 ausgewählten Arbeiten gibt es nur 35 Artikel, in denen man die Phasendifferenz mit Hilfe sta-

Tabelle 4-6: Die Prozentsätze der Phasenpaare, die das Kriterium des Nicht-Überschneidens erreichen

Zeitschriften	JABA										Summe
Jahrgänge	'68	'69	'70	'71	'72	'73	'74	'75	'76	'70-'76	
Gesamtzahl der Phasenpaare	399	262	462	423	577	486	672	569	916	489	292
											759
											6306
Sich nicht-überschneidende Phasenpaare	137	90	135	111	148	141	241	210	319	161	116
											243
											2052
% ¹⁾	34.3	34.4	29.2	26.2	25.6	29.0	35.9	36.9	34.8	32.9	39.7
											32.0
											32.5

$$1) \% = \frac{\text{Sich nicht-überschneidende Phasenpaare}}{\text{Gesamtzahl der Phasenpaare}} \cdot 100$$

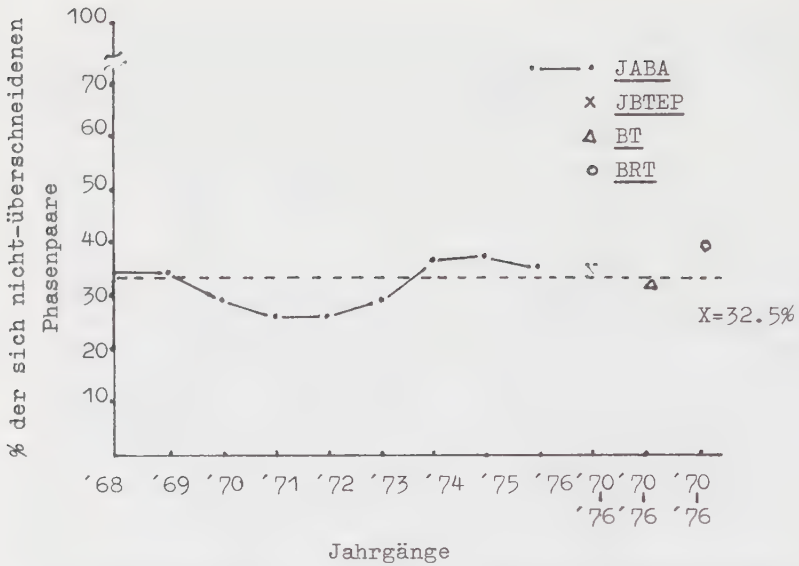


Abb. 4-2: Prozentsätze der sich nicht-überschneidene Phasenpaare

tistischer Prüfverfahren kontrolliert. Die Tabelle 4-7 gibt eine Übersicht über die verwendeten Verfahren. In den übrigen Artikeln werden die Forschungsergebnisse nur durch deskriptive Statistik dargestellt, d.h. durch Angaben über Zunahme oder Abnahme des Mittelwertes von einer Phase zur anderen, oder durch intuitive Beurteilung des Anstiegs oder Abstiegs im Kurvenverlauf. Die Angemessenheit der benutzten statistischen Prüfverfahren wird im Abschnitt 4.5 der vorliegenden Arbeit eingehend erörtert.

Tabelle 4-7: Übersicht über die verwendeten statistischen Prüfverfahren für die experimentellen Einzel-falluntersuchungen in den vier ausgewählten Zeitschriften zur Verhaltensmodifikation

Anzahl der Vpn	Die verwendeten statistischen Prüfverfahren	Die Artikel, deren Autoren das Prüfverfahren verwenden
N=1	t-Test für verbundene Stichproben (related samples)	Brady, et al.(1974) Leitenberg, et al. (1970)
	t-Test	Lanyon et al.(1976) Kubany et al.(1971)
	H-Test(Ein-Weg-Rangvarianzanalyse nach Kruskal und Wallis)	Calhoun & Matherne (1975)
	Varianzanalyse nach Gentile et al.(in <u>JABA</u> , 1972, 193-198)	Epstein et al. (1974)
	Zeitreihenanalyse	Schnelle et al.(1975)
	Chi-Quadrat-Test	Rubin & Stolz(1974)
	Mann-Whitney U-Test	Calhoun & Matherne (1975). Arnold et al.(1976)
N=1-Gruppe	Vorzeichen-Test	Sachs et al.(1972)
	t-Test für gepaarte Stichproben (matched samples)	Renne & Creer, 1976, (N=4). Glover & Gary, 1976, (N=8).
	t-Test für korrelierte Stichproben	Miller & Weaver, 1976, (N=9). Eriksson et al., 1975, (N=25).
	t-Test	Jones & Eimers, 1975, (N=28). O'Leary et al., 1969, (N=7).

Tabelle 4-7 (Fortsetzung)

Anzahl der Vpn	Die verwendeten statistischen Prüfverfahren	Die Artikel, deren Autoren das Prüf- verfahren verwenden
N=1- Gruppe	t-Test	Hollander & Plutchik, 1972, (N=13)
	Varianzanalyse für wiederholte Messungen	O'Leary et al., 1969, (N=7). Kendall et al., 1975, (N=30). Epstein et al., 1976, (N=6). Creer & Davis, 1975, (N=9). Drabman & Lahey, 1974, (N=1 Klasse). Glynn & Thomas, 1974, (N=9). Kazdin & Klock, 1973, (N=12). Feingold & Mahoney, 1975, (N=5).
	Zufallsblöcke-Varianzanalyse	Richard & Aaunders Jr., 1971, (N=8).
	Varianzanalyse	Thomas et al., 1968, (N=28). Boudewyns, 1976, (N=40).
	Korrelation zwischen unabhän- gigen und abhängigen Variablen	Kazdin & Klock, 1973, (N= 1 Lehrer und 12 Schüler).
	Wilcoxon's Rang-Vorzeichen-Test	Winkler, 1970, (N=66). O'Brien et al., 1969, (N=13).

Tabelle 4-7 (Fortsetzung)

Anzahl der Vpn	Die verwendeten statistischen Prüfverfahren	Die Artikel, deren Autoren das Prüf- verfahren verwenden
N=4- Gruppe	McNemar-Test(Chi-Quadrat-Test)	Kaufman & O'Leary, 1972, (N=16).
	Chi-Quadrat-Test	McInnis et al., 1974, (N=56). Curran et al., 1973, (N=9).
	Omega-Quadrat Schätzung nach Hays(Hays, 1969, S.327).	Glynn & Thomas, 1974, (N=9).
	Mann-Whitney-Test	Doty et al., 1974, (N=28).
	Randomisierung-Test für Paral- lele Paare	Barlow et al. 1972, (N=4). Hauserman et al., 1973, (N=7).

- (2) Wie groß ist die Anzahl der Beobachtungen pro Phase (d.h. die Phasenlänge), und ist sie für die experimentelle Einzelfalluntersuchung angemessen? Barlow & Hersen (1973) behaupten, daß es mindestens drei Beobachtungen in jeder Phase sein müssen. Ist das genug? Wie haben es die meisten Forscher gemacht?

Tabelle 4-8 und 4-9 geben an, wie viele Beobachtungen pro Phase die Forscher durchgeführt haben. Bei jedem Artikel wurde jede gebrauchte Phasenlänge höchstens nur einmal aufgezeichnet. Z.B. bei einem Artikel mit dem ABAB-Design waren die Phasenlängen jeweils 6-8-5-5. D.h., in der ersten Baselinephase gab es 6 Beobachtungspunkte, in der ersten Behandlungsphase 8 Beobachtungspunkte, in der zweiten Baselinephase 5 und in der zweiten Behandlungsphase auch 5. In diesem Fall wurde je eine Häufigkeit für die Phasenlänge 6, 8 und 5 aufge-

Tabelle 4-8: Die Auftretenshäufigkeit der verschiedenen Phasenlängen in JABA (1968-1976)

Phasen- länge	Häufig- keit	Phasen- länge	Häufig- keit	Phasen- länge	Häufig- keit	Phasen- länge	Häufig- keit
1	35	17	37	33	5	51	3
2	69	18	31	34	5	52	1
3	96	19	25	35	8	54	1
4	122	20	33	36	10	55	1
5	123	21	23	37	3	58	2
6	125	22	13	38	4	64	1
7	106	23	18	39	2	65	1
8	106	24	16	40	3	67	1
9	88	25	10	41	2	72	1
10	102	26	8	42	2	76	2
11	60	27	7	43	4	83	1
12	56	28	11	44	1	91	1
13	42	29	5	45	4	113	1
14	50	30	13	46	1	120	1
15	48	31	7	49	1	125	1
16	34	32	6	50	2	177	1

Tabelle 4-9: Die Auftretenshäufigkeit der verschiedenen Phasenlängen in JBTEP, BT und BRT zusammen (1970-1976)

Phasen- länge	Häufig- keit	Phasen- länge	Häufig- keit	Phasen- länge	Häufig- keit	Phasen- länge	Häufig- keit
1	21	11	17	21	1	35	1
2	27	12	17	22	3	40	1
3	55	13	11	23	1	42	1
4	50	14	9	24	2	43	1
5	58	15	13	25	3	53	1
6	45	16	4	26	1		
7	35	17	3	28	3		
8	43	18	9	29	1		
9	30	19	6	30	3		
10	25	20	9	32	1		

Auftretenshäufigkeiten

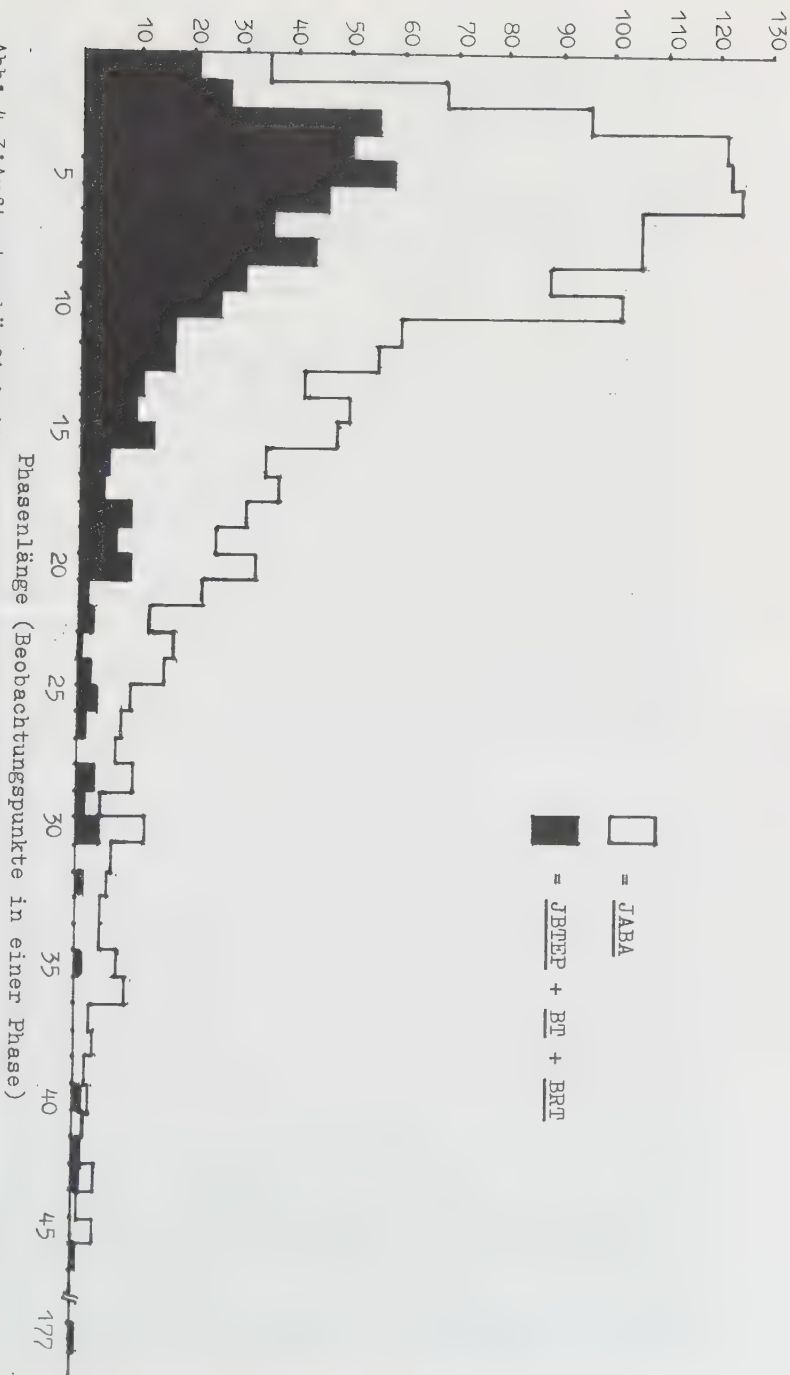


Abb: 4-3: Auftretenshäufigkeiten jeder Phasenlänge in JABA und den anderen drei Zeitschriften

zeichnet, anstatt eine für 6, eine für 8 und zwei für 5. Die Beobachtungspunkte in der Follow-up-Phase wurde nicht berücksichtigt.

Die Verteilungen der Frequenzen jeder Phasenlänge in den Artikeln aus JABA und den anderen drei Zeitschriften(JBTEP, BT und BRT zusammen) werden in Abb. 4-3 wiedergegeben. Aus Abb. 4-3 ist ersichtlich, daß die Formen der beiden Verteilungen sehr ähnlich sind. Sie sind beide linksschief und verlaufen fast parallel. Um den Übereinstimmungsgrad beider Verteilungen zu bestimmen, wurde die Korrelation zwischen den Auftretensfrequenzen der einzelnen Phasenlängen beider Verteilungen berechnet. Es ergibt sich eine hoch signifikante Korrelation ($r=0.95$; $N=53$). In die Berechnung wurden die Phasenlängen von 54 bis 177 der Verteilung JABA nicht mit einbezogen, weil es bei der Verteilung der anderen drei Zeitschriften keine Phasenlängen über 54 Beobachtungspunkte gibt. Aus den Verteilungen ist ersichtlich, daß in den meisten Untersuchungen nur kurze Phasen vorliegen, d.h. die Mehrheit der Autoren erhob weniger als 30 Datenpunkte in einer Phase. Wenn man die Beobachtungen in einer Phase als eine Stichprobe aus der Population der Verhaltensweisen in jener Phase betrachtet, dann handelt es sich meistens um eine kleine Stichprobe.

Dieses Phänomen spiegelt vielleicht, außer den zeitlichen und ethischen Gründen, die Berücksichtigung wider, daß die lange Phasenlänge genug Zeit für den Aufbau der neu konditionierten Verstärker einräumt. Wenn dies geschieht, dann hat wegen der Wirkung der neu gewonnenen Verstärker die Absetzung der experimentellen Behandlung in der nächsten Experimentalphase einer Studie keine Abnahme der Verhaltensrate zur Folge.

- (3) Wie hoch soll die Beobachtungsübereinstimmung zwischen den Beobachtern sein? Tunner & Florin (1969) fordern eine Beobachterreliabilität von mindestens $r=0.8$. Reicht dieses Kriterium aus? Wie haben es die meisten Forscher gemacht?

Um die Zuverlässigkeit (Reliabilität) der Verhaltens-

beobachtung zu sichern, ist es erforderlich, daß zwei oder mehr Beobachter unabhängig voneinander nach zuvor festgelegten Kodierungen und Kriterien die Verhaltensweisen aufzeichnen. Die Zuverlässigkeit der Verhaltensbeobachtung wird durch die Reliabilitätskoeffizienten numerisch ausgedrückt. Normalerweise wird das Reliabilitätsmaß der Beobachtungen durch folgende Formel ausgerechnet:

$$\text{Prozentsatz der Übereinstimmung} = \frac{\text{Die Summe aller Übereinstimmungen}}{\text{Die Summe aller Übereinstimmungen} + \text{Die Summe aller Nicht-Übereinstimmungen}} \cdot 100$$

Diese Formel ist für die Art der Datenerhebung angemessen, bei der die Verhaltensweise unter dem Aspekt Auftreten oder Nicht Auftreten aufgezeichnet wird. Aber wenn die Dauer einer Verhaltensweise registriert wird, drückt man die Übereinstimmung der beiden Beobachter durch den Korrelationskoeffizienten aus. Normalerweise rechnet man bei jeder Vp für jede Verhaltensweise den **Prozentsatz** oder den Korrelationskoeffizienten der Übereinstimmung aus; wenn mehrere Vpn und/oder mehrere Verhaltensweisen untersucht werden, kann es deshalb mehrere Übereinstimmungskoeffizienten geben, aus denen der Mittelwert berechnet wird. In den untersuchten Zeitschriften sind nicht in allen ausgewählten Artikeln die Mittelwerte der Übereinstimmungsprozentsätze angegeben. Tabelle 4-10 zeigt, daß bei ca. 70% der ausgewählten Artikel von JABA, aber nur bei ca. 35% der anderen drei Zeitschriften (JBTEP, BT und BRT) die durchschnittlichen Übereinstimmungsprozentsätze genannt wurden. Bei einem Teil der übrigen Artikel wurde eine Instrumentalmessung vorgenommen, die objektiv ist und bei der sich deshalb ein Übereinstimmungsprüfung erübrigt. Bei einigen Artikeln wurden überhaupt keine Abgabe über die Übereinstimmung zwischen den Beobachter gemacht, bei anderen wurden nur unvollständige Mitteilungen gegeben,

z.B., daß die Spannweite der Übereinstimmungsprozentsätze von 80% bis 99% gehe. Wenn man angibt, die Spannweite der Übereinstimmungsprozentsätze gehe von 80% bis 99% mit einem Median von 92%, dann wird dieser Median auch als Mittelwert betrachtet, weil er die gleiche Funktion wie der Mittelwert ausübt, nämlich die Übereinstimmungsprozentsätze zu repräsentieren. Aus den ausgewählten Artikeln, die die durchschnittlichen Übereinstimmungsprozentsätze angeben haben, ergeben sich die Daten von Tabelle 4-11. Es ist bemerkenswert, daß der durchschnittliche Übereinstimmungsprozentsatz jedes Jahrgangs oder jeder Zeitschrift ungefähr 93% beträgt. Abb. 4-4 verdeutlicht, daß dieser Trend sehr stabil ist ($\sigma=0.85\%$). Dieser durchschnittliche Prozentsatz der Beobachtungsübereinstimmung (93.6%) sollte als ein Kriterium für das Beobachtertraining betrachtet werden.

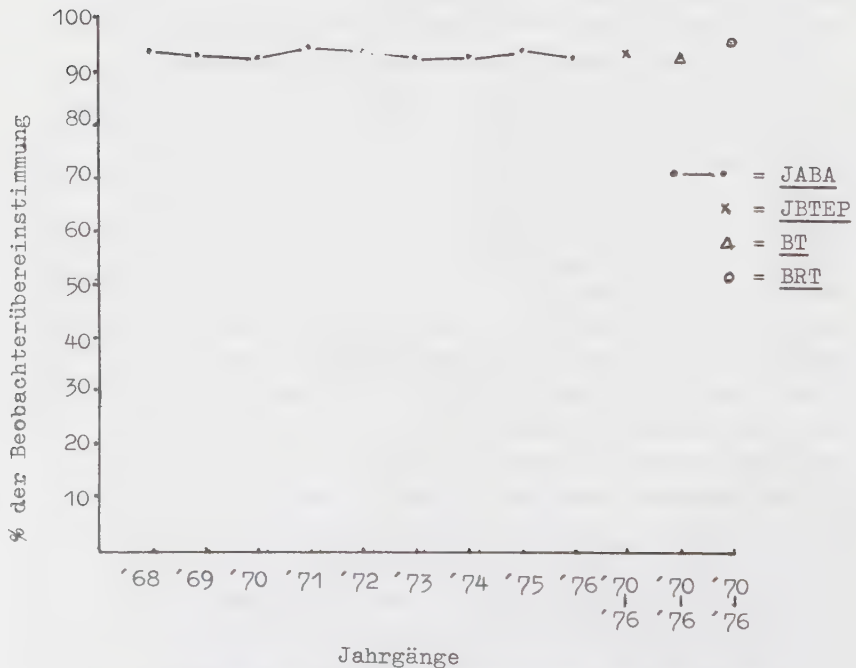


Abb. 4-4: Durchschnittliche Prozentsätze der Beobachter-Übereinstimmung

Tabelle 4-10: Prozentsätze der Artikel, die die durchschnittliche Übereinstimmung zwischen Beobachtern angegeben haben

Zeit- schriften	<u>JABA</u>										<u>JBTEP</u>	<u>BT</u>	<u>BRT</u>
	'68	'69	'70	'71	'72	'73	'74	'75	'76	'70-'76	'70-'76	'70-'76	'70-'76
Jahrgänge	'68	'69	'70	'71	'72	'73	'74	'75	'76	'70-'76	'70-'76	'70-'76	'70-'76
%	57.1	58.8	61.5	50.0	84.6	80.0	76.2	84.6	76.5	27.3	44.9	33.3	
Mittel- wert	69.9										35.2		

Tabelle 4-11: Die durchschnittlichen Prozentsätze der Beobachterübereinstimmung

Zeit- schriften	<u>JABA</u>										<u>JBTEP</u>	<u>BT</u>	<u>BRT</u>	Mittelwert der vier Zeitschrif- ten
	'68	'69	'70	'71	'72	'73	'74	'75	'76	'70-'76	'70-'76	'70-'76	'70-'76	
Jahrgänge	'68	'69	'70	'71	'72	'73	'74	'75	'76	'70-'76	'70-'76	'70-'76	'70-'76	
%	93.8	93.4	93.3	94.8	93.6	92.7	92.6	93.9	93.3	93.7	92.8	95.5		93.6

- (b) Zur Fragestellung hinsichtlich der externen Validität:
Wenn man die experimentelle Einzelfalluntersuchung als
eine Forschungsintervention betrachtet, wie weit ist sie
im Hinblick auf Versuchspersonen, Experimentatoren,
Situationen, unabhängige Variablen und abhängige Varia-
blen verallgemeinert worden?
- 1) Zunächst wird berichtet, wie häufig die verschiedenen Versuchspläne der experimentellen Einzelfalluntersuchung Anwendung gefunden haben. Die Versuchspläne sind in sechs Gruppen klassifiziert worden (siehe Abschnitt 3.2.2 der vorliegenden Arbeit). Um die verschiedenen Versuchspläne gegeneinander abzugrenzen, werden folgende operationale Definitionen verwendet:
1. Das ABAB-Design: Die Versuchspläne, die einen Bestandteil des Reversionsplans (d.h. ABA-Design) enthalten, werden dieser Kategorie zugeordnet, z.B. A-B-A-B; A-B-C-B; A-B-C-A-C; A-BC-A-BC; A-B-BC-BD-BC; A-B-A-C-D-C; usw. Ein Treatmentpaket, das zwei oder mehr Treatments enthält, wird als ein Treatment betrachtet.
 2. Das ABACA-Design: Versuchspläne, mit denen man zwei oder mehr experimentelle Behandlungen gegenüber einer Baseline überprüft, gehören zu dieser Kategorie.
 3. Das ABABB'B"-Design: Bedingung für diese Kategorie ist, daß die Wirkung der Behandlung zunächst gegenüber der Baseline experimentell überprüft und dann die Behandlung variiert wird.
 4. Das BC-C-BC-Design: Der Unterschied zwischen dem ersten und vierten Design besteht darin, daß beim ersten Design die Behandlung gegenüber der Baseline überprüft wird, während beim vierten ein Behandlungspaket gegenüber einem seiner Elemente untersucht wird. Es ist der Hauptzweck dieses Designs, die Wirkung der einzelnen Elemente des Behandlungspakets zu analysieren, z.B. A-BC-B-BC; BCD-BC-BCD; B-BC-B.
 5. Das Multiple-Baseline-Design: Bei diesem Design erstellt man zunächst gleichzeitig zwei oder mehr Baselines von verschiedenen Verhaltensweisen einer Versuchsperson (oder Baselines von einer Verhaltensweise einer Vp in verschiedenen Situationen, oder Baselines einer Ver-

haltensweise bei verschiedenen Vpn), dann setzt man die Behandlung zu verschiedenen Zeitpunkten für verschiedene Baselines ein. Z.B.

$$\begin{cases} A-B-A-B \\ A-A-B-B \\ A-A-A-B; \end{cases} \begin{cases} A-B-A-A \\ A-A-B-A \\ A-A-A-B; \end{cases} \begin{cases} A-B-C-A \\ A-A-B-C; \\ \end{cases} \begin{cases} A-B-BC-A-A \\ A-B-A-BC-A \\ A-B-A-A-BC \end{cases}$$

6. Das Multiple-Kontingenz-Design: Bei diesem Versuchsplan wird ein und dasselbe Verhalten einer Vp unter verschiedenen Stimulusbedingungen unterschiedlich behandelt. solche Stimulusbedingungen können in unterschiedlichen Tageszeiten (z.B. vor- vs. nachmittags) unterschiedlichen Situationen (z.B. Schule vs. Haus) oder unterschiedlichen Versuchsleitern etc. bestehen. Voraussetzung ist, daß der Vp Gelegenheit zum Stimulus-Diskriminationslernen geboten wird. Wenn Treatment A am Vormittag verabreicht wird, muß Treatment B am Nachmittag gegeben werden, oder umgekehrt. Z.B.

$$\begin{cases} A-B-A-B \\ B-A-B-A; \end{cases} \begin{cases} A-B-C-B \\ A-C-B-C. \end{cases}$$

Tabelle 4-12 gibt die Anwendungshäufigkeit der sechs Versuchspläne in den vier ausgewählten Zeitschriften an. Es ist bemerkenswert, daß die erste Versuchsanordnung (ABAB-Design) und die fünfte (das Multiple-Baseline-Design) am stärksten eingesetzt wurden. Abb. 4-5 verdeutlicht dieses Phänomen. Dieser Unterschied erweist sich bei einem Chi-Quadrat-Test als sehr signifikant ($\chi^2=543.69$, Fg.=5, $p<0.01$).

Tabelle 4-12: Anwendungshäufigkeit der verschiedenen Einzelfall-Experimentaldesigns
in den vier ausgewählten Zeitschriften¹⁾

Zeit- schriften	<u>JABA</u>												<u>TJREP</u>		<u>BT</u>		<u>BRT</u>		Σ	%
	'68	'69	'70	'71	'72	'73	'74	'75	'76	'70-'76	'70-'76	'70-'76	'70-'76	'70-'76	'70-'76	'70-'76	'70-'76	'70-'76		
Jahrgänge																				
ABAB	18	11	16	9	17	19	25	4	15	35		25						12	206	53.5
ABACA	0	0	0	0	0	4	2	2	2	2		2						2	16	4.2
ABABB'B"	1	0	2	2	0	0	0	0	1	1		2						0	9	2.3
B-BC-B	1	1	2	2	3	2	0	1	0	3		4						1	20	5.2
{ ABB AAB ABAB BABA	1	4	7	5	6	11	22	20	20	14		15						3	128	33.2
	0	1	1	0	1	1	0	0		1		1						0	6	1.6
Σ	21	17	28	18	27	37	49	27	40	56		49						18	335	100.0

1) Anhang 7 gibt die Seiten der Artikel an, in denen die verschiedenen Einzelfall-Experimentaldesigns angewendet wurden.

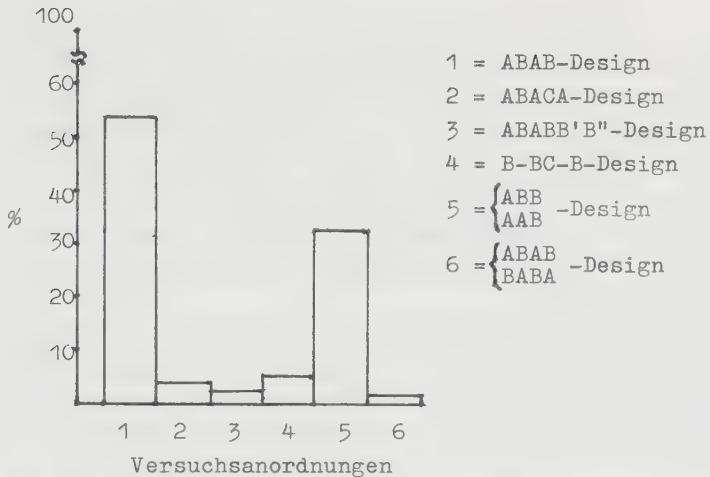


Abb.4-5:Prozentuelle Verteilung der verschiedenen Einzelfall-
 Experimentaldesigns in den vier ausgewählten Zeit-
 schriften

2) Wie ist die experimentelle Einzelfalluntersuchung hin-
sichtlich der Versuchspersonen verallgemeinert worden?

Es geht um die Frage, ob die experimentelle Einzelfall-
 untersuchung für alle Altersgruppen anwendbar ist.

M.a.W. ist die experimentelle Einzelfalluntersuchung
 nur für Behinderte bzw. nur für kleine Kinder geeignet?
 In der vorliegenden Untersuchung werden die Vpn in zwei
 Kategorien eingeteilt: Behinderte und Nichtbehinderte.
 In jeder dieser Kategorien wird eine weitere Unterteil-
 ung der Vpn nach fünf Altersgruppen vorgenommen:

1. Klein- und Vorschulkinder bzw. Kinder bis 6 Jahre.
2. Schüler der Elementarschule (1.-6. Schulklasse) bzw.
 Kinder des entsprechenden Alters (7-12 Jahre).
3. Schüler der höheren Schule (7.-12. Schulklasse) bzw.
 Vpn des entsprechenden Alters (13-18 Jahren).
4. Studenten der Hochschule bzw. Vpn des entsprechenden
 Alters (19-22 Jahre).
5. Erwachsene (ab 23 Jahre).

Der Begriff "behindert" (retarded) wird wie folgt
 operational definiert: Diejenigen, die wegen körper-

licher, intellektueller oder psychischer Behinderung nicht die normale Schule besuchen können, und diejenigen, die eine psychotische Krankheit haben, werden als "Behinderte" betrachtet, z.B. Schüler in Sonderschulen (Blinde, Sehbehinderte, Gehörlose, Schwerhörige, Geistig behinderte, Körperbehinderte, Sprachbehinderte, Lernbehinderte, Verhaltensgestörte) und psychisch Kranke (Vpn mit Schizophrenie, Autismus, Mongolismus, Hirnverletzungen, usw.).

Vpn mit bestimmten Phobien (z.B. Schulangst, Agoraphobie, Messerphobie, Claustrophobie) werden nicht als Behinderte bezeichnet, weil solche Phobien nur im Zusammenhang mit bestimmten Gelegenheiten oder Situationen auftreten, während sich die betreffenden Personen in anderen Situationen normal verhalten.

Die Klassifikation von Behinderten und Nichtbehinderten ist dichotomisch. Unter Nichtbehinderten werden normale Schulkinder, Studenten, Berufstätige usw. verstanden.

Die Verteilung der verschiedenen Altersgruppen in JABA und den anderen drei Zeitschriften ist in Tabelle 4-13 zusammengestellt. Um eine Inflation der Häufigkeit zu vermeiden, wurde jede Altersgruppe bei jedem Artikel nur einmal aufgezeichnet. Z.B. bei einer Beschreibung "9 behinderte Vpn im Alter von 12 bis 15 Jahren wurden dem Experiment unterzogen" wurde einmal die Kategorie "behindert: 7-12 Jahre" und einmal die Kategorie "behindert: 13-18 Jahre" registriert.

In JABA ist die Anzahl der nichtbehinderten Vpn höher als die Anzahl der behinderten ($\chi^2=12.8$, Fg.=1, $p<0.01$), aber dieser Unterschied liegt bei den anderen drei Zeitschriften nicht vor (zum Teil ist er umgekehrt) ($\chi^2=0.31$, Fg.=1, nicht signifikant). Es gibt kaum konstante Gesetzmäßigkeiten über die Häufigkeitsverteilung der verschiedenen Altersgruppen. Aber es zeigt sich, daß die experimentelle Einzelfalluntersuchung für alle Altersgruppen und sowohl für Nichtbehinderte als auch für Behinderte anwendbar ist.

Tabelle 4-13: Verteilung der Altersgruppen der Vpn in den vier ausgewählten Zeitschriften

Zeit- schriften	<u>JABA</u>											<u>JBMEP</u>	<u>BT</u>	<u>BRT</u>
	Jahrgänge	'68	'69	'70	'71	'72	'73	'74	'75	'76	Σ	'70-'76	'70-'76	'70-'76
bis 6 Jahre		6	2	6	5	6	2	10	4	5	46	1	0	3
7-12 Jahre		7	8	9	4	8	12	14	5	7	74	3	0	12
13-18 Jahre		2	0	4	3	2	7	7	4	4	33	2	2	6
19-22 Jahre		1	1	1	1	1	3	3	1	6	18	4	3	5
ab 23 Jahre		1	0	1	0	0	4	5	7	3	21	13	8	13
Σ		192										75		
bis 6 Jahre		1	0	2	3	5	3	4	3	5	26	3	0	2
7-12 Jahre		4	4	7	3	5	8	7	4	8	50	14	0	9
13-18 Jahre		0	3	4	3	1	3	3	4	1	22	6	2	5
19-22 Jahre		0	0	3	0	1	2	0	1	4	11	4	3	4
ab 23 Jahre		1	3	4	1	1	3	2	0	4	19	11	8	11
Σ		128										82		

3) Wie ist die experimentelle Einzelfalluntersuchung hinsichtlich des Experimentators verallgemeinert worden?

Bei den in natürlichen Situationen durchgeführten Experimenten ist es häufig erforderlich, daß die Rolle des Experimentators von Laien übernommen wird. Können die Laien bei der experimentellen Einzelfalluntersuchung als Versuchsleiter herangezogen werden?

In der vorliegenden Untersuchung wird der Begriff "Laien" als Gegenteil zu Experten der Psychologie, oder noch präziser, Experten der Verhaltensmodifikation verstanden. Der Kategorie "Experten" werden die Verhaltenstherapeuten, die "Eltern" der hausartigen Rehabilitationsanstalt (achievement place), der Autor eines Artikels (Forscher), ausgebildete Assistenten oder Studenten im Bereich Psychologie zugeordnet.

Die Kategorie "Laien" wird in vier Subkategorien gegliedert:

1. Lehrer: Professor (mit Ausnahme des Verfassers selbst), Lehrer in Schule, Kindergarten oder Sonderschule.
2. Eltern: Vater und/oder Mutter der Versuchsperson.
3. Peers: Teaching Practors (Studenten, die dem Professor bei der Lehrtätigkeit helfen), Peers, Geschwister, und Mitschüler.
4. Personal: Pfleger oder Krankenschwester im Krankenhaus, Personal in Institut, Betrieb oder Schule.

Die Auftretenshäufigkeit der Experten, Lehrer, Eltern, Peers und des Personals als Versuchsleiter ist für JABA und die anderen Zeitschriften in Tabelle 4-14 aufgeführt. Bei den drei Zeitschriften (JBTEP, BT und BRT) wurden als Versuchsleiter mehr Experten (59.5%) als Laien (40.5%) eingesetzt. (Mit Chi-Quadrat-Test, $\chi^2=4.57$, Fg.=1, $p<0.05$). In JABA ist das Verhältnis umgekehrt, d.h. bei den Untersuchungen nahmen als Versuchsleiter mehr Laien (57.5%) als Experten (42.5%) teil. Diese prozentuelle Differenz wird auch mit Hilfe des Chi-Quadrat-Tests als signifikant bestätigt ($\chi^2=5.69$, Fg.=1, $p<0.02$). In JABA sind 38.2% der Versuchsleiter Lehrer, in den anderen drei Zeitschriften nur 15.1%.

Tabelle 4-14: Auftretenshäufigkeit der Experten, Lehrer, Eltern, Peers und des Personals als Versuchsleiter in den vier ausgewählten Zeitschriften

Zeit- schriften	<u>JABA</u>										Σ	%	<u>JBTEP</u> <u>BKT</u> <u>BT</u>			Σ	%
	'68	'69	'70	'71	'72	'73	'74	'75	'76	'70-'76			'70-'76	'70-'76			
Jahrgänge	11	3	9	9	6	13	20	13	19	108	42.5	35	15	25	75	59.5	
Experten	9	8	12	7	17	16	14	10	4	97	38.2	8	0	11	19	15.1	
Lehrer	1	3	1	1	3	3	2	1	0	15	5.9	3	0	4	7	5.5	
Eltern	0	1	0	1	0	1	2	0	1	6	2.4	1	0	3	4	3.2	
Peers	0	3	5	0	1	1	5	3	10	28	11.0	9	3	9	21	16.7	
Personal																	

Obwohl in allen vier Zeitschriften nur selten der Einsatz von Eltern und Peers als Versuchsleiter erwähnt wird, zeigen uns die wenigen Fälle doch, daß sie sich nach kurzem Training durchaus als Versuchsleiter eignen können.

4.) Wie ist die experimentelle Einzelfalluntersuchung hinsichtlich der Situation verallgemeinert worden?

Es taucht nun die Hauptfrage auf, ob die experimentelle Einzelfalluntersuchung in der natürlichen Situation anwendbar ist.

In der vorliegenden Untersuchung werden die Situationen, in denen man Untersuchungen mit der experimentellen Einzelfalluntersuchung durchführte, nach sieben Kriterien unterschieden.

1. Labor: Hierunter versteht man eine Forschungsstätte, die von anderen Räumen isoliert, schalldicht und spezifisch eingerichtet ist, in der die Störfaktoren so weit wie möglich ausgeschaltet werden können, z.B. Versuchsräume in klinischen Forschungszentren, Kinderforschungszentren, Universitäten, Krankenhäuser, usw.

2. Schule: Wenn die Untersuchung im Klassenzimmer oder auf dem Spielplatz irgendeiner Art von Schule (von Kindergarten bis Hochschule einschließlich der Sonderschule) durchgeführt wurde, wird diese Situation als "Schule" bezeichnet.

3. Klinik: Wenn die Untersuchung in einer Station eines Krankenhauses, besonders eines psychiatrischen Krankenhauses, stattfand, wird sie unter "Klinik" aufgezeichnet.

4. Haus: Bei dieser Situation wurde die Untersuchung im Haus oder dessen Umgebung (z.B. Spielplatz) der Vp durchgeführt. Ein Studentenwohnheim läßt sich auch dieser Kategorie zurechnen.

5. Rehabilitationsanstalt für Kriminelle: Dies ist eine Anstalt, in der Jugendliche oder erwachsene Delinquenten gefangen sind und eventuell durch Berufsausbildung später wieder in die Gesellschaft eingegliedert werden, z.B. Home for Delinquents, Institution for Offender, Security Correctional Institution for Adult Felons,

Tabelle:4-15 Überblick über die Frequenzverteilung der verschiedenen Situationen, in denen Untersuchungen mit einer experimentellen Einzelfalluntersuchung durchgeführt wurden

Zeit- schriften	<u>JABA</u>											<u>JBRP</u>	<u>BR</u>	<u>BRP</u>	Σ	%
	'68	'69	'70	'71	'72	'73	'74	'75	'76	Σ	%					
Jahrgänge	'68	'69	'70	'71	'72	'73	'74	'75	'76	Σ	%	'70-'76	'70-'76	'70-'76	Σ	%
Labor	4	2	0	1	0	5	4	1	5	22	8.7	9	8	12	29	23.6
Schule	12	10	16	9	16	18	20	16	12	129	51.0	15	0	13	28	22.8
Klinik	2	4	8	5	6	10	10	5	5	55	21.7	23	8	12	43	34.9
Haus	1	2	1	1	3	0	3	1	3	15	5.9	6	1	7	14	11.4
Rehabili- tations- anstalt	1	0	1	2	1	2	0	1	5	13	5.1	1	0	2	3	2.4
Sonstige	1	0	1	0	0	3	7	3	4	19	7.5	2	1	3	6	4.9

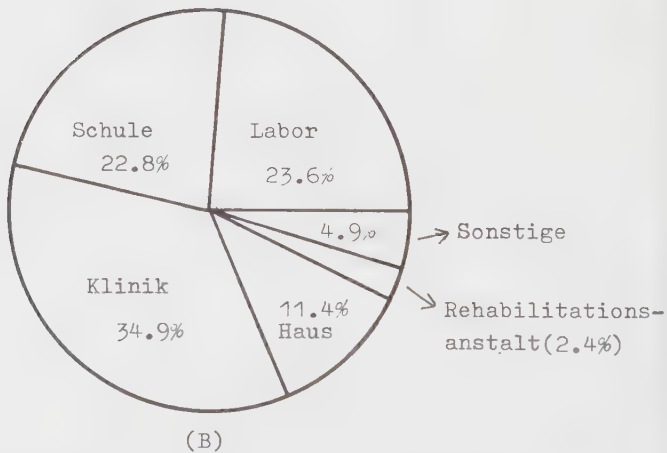
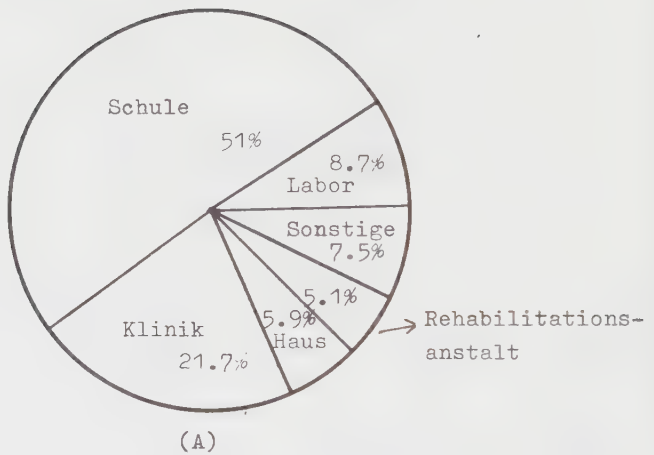


Abb. 4-6: Prozentuelle Verteilung der verschiedenen Situationen, in denen Untersuchungen mit einer experimentellen Einzelfalluntersuchung durchgeführt wurden, in JABA (A) und in den anderen drei Zeitschriften (B).

Residential Treatment for Predelinquents, Achievement Place. Achievement Place ist eine kommunale hausartige Rehabilitationsanstalt für Kriminelle oder kriminell orientierte Jugendliche, in der "Eltern" eingesetzt werden, die die Techniken der Verhaltensmodifikation beherrschen.

6. Sonstige: Diese Kategorie faßt alle Situationen, die den obengenannten Kategorien nicht zugeordnet werden können, z.B. auf der Straße, im Schulbus, Schwimmbad, Mensa, Sitzungssaal, Wohnviertel, Säuglingsheim, Firma, Jugendzentrum, Sozialamt und bei der Freizeitgestaltung.

Tabelle 4-15 gibt einen Überblick über die Frequenzverteilung der verschiedenen Situationen, in denen Untersuchungen mit der experimentellen Einzelfalluntersuchung durchgeführt wurden. In Abb. 4-6 ist ersichtlich, daß in JABA nur 8.7% und in den anderen Zeitschriften 23.6% der Untersuchungen mit einer experimentellen Einzelfalluntersuchung in Laborsituation durchgeführt wurden, und daß mehr als die Hälfte der in JABA dargestellten Untersuchungen in der Schule stattgefunden haben. Dies weist darauf hin, daß die experimentelle Einzelfalluntersuchung in natürlichen Situationen durchführbar ist.

5) Wie ist die experimentelle Einzelfalluntersuchung hinsichtlich der unabhängigen Variablen verallgemeinert worden?

In der vorliegenden Untersuchung werden die unabhängigen Variablen der ausgewählten Artikel nach folgenden Kategorien klassifiziert:

1. Primäre Verstärkung: Die erwünschte Verhaltensweise wird kontingent mit einem unkonditionierten Reiz verstärkt, z.B. mit Speisen, Süßigkeiten, Zigaretten oder anderen Materialien als Verstärker.

2. Soziale Verstärkung: Die erwünschte Verhaltensweise wird kontingent mit Zuwendung, verbalem Lob, oder physischem Kontakt verstärkt.

3. Tokensystem: Die erwünschte Verhaltensweise wird kontingent mit stellvertretenden Verstärkern (Token)

verstärkt, die später in einem bestimmten, vorher festgesetzten Verhältnis gegen tatsächliche Verstärker eingetauscht werden können.

4. Aktivitätskonsequenz: Das Zielverhalten wird kontingent mit Aktivität, die die Wirkung von positiver oder negativer Verstärkung hat, verstärkt, z.B. kontingent mit Vorrecht für extra Aktivität (Spiel, Fernseh-Gucken, Radio-Hören, Geschichten-Hören, Freizeit, Urlaub, frühzeitigem Feierabend, usw.); mit dem Premark-Prinzip (mit der häufig vorkommenden Aktivität der Versuchsperson als Verstärker); kontingent mit tutorialen Nachhilfen nach der Schulzeit bei der Schulnote D oder F (d.h. der Schüler mußte nach der Schulzeit noch zum Nachhelfen bleiben, wenn er die Schulnote D oder F bekommen hat; Die Kinder mußten die vorausgehende Aufgabe erledigen, ehe ihnen andere Aktivität erlaubt wurde; Schlafenszeit-Konsequenz (wenn das Kind 10 Minuten weniger für die Hausaufgabe gearbeitet hatte, mußte es 10 Minuten früher ins Bett gehen).

5. Feedback: Die Versuchsperson wird über ihre eignen Reaktionen informiert. Selbst-Aufzeichnung wird auch als eine Art von Feedback betrachtet. Biofeedback gehört auch zu dieser Kategorie.

6. Selbstkontrolle: Die erwünschte (oder unerwünschte) Verhaltensweise wird dadurch verstärkt (oder abgeschwächt), daß die Versuchsperson ihr zu änderndes Verhalten selbst diagnostiziert (self assessment), selbst aufzeichnet (self recording) und selbst über die Kontingenz entscheidet (self administration of reinforcement).

7. Training: Eine Reihe von Maßnahmen für den Aufbau einer neuen Verhaltensweise kennzeichnet dieses Verfahren. Es enthält Instruktion, Demonstration, Imitation oder Übung, und dann Feedback oder Verstärkung.

8. Verstärkerentzug: Die unerwünschte Verhaltensweise wird durch kontingenten Entzug oder Vorenthalten des Verstärkers oder Tokens abgeschwächt, z.B. kontingent mit der Tokenpunkteeinbuße (token fine).

9. Ignorieren: Die unerwünschte Verhaltensweise wird dadurch abgeschwächt, daß die Zuwendung oder Aufmerk-

samkeit vorenthalten wird, die zuvor diese Verhaltensweise verstärkt hat. Z.B. wird der Pfleger angewiesen, sich von dem Patienten abzuwenden, sobald der Patient eine bizarre Verbalisation machte.

10. Darbietung aversiver Reize: Die unerwünschte Verhaltensweise wird kontingent mit einem aversiven Reiz vermindert oder unterdrückt, z.B. Schock, Ammoniak, saurer Zitronensaft, verbale oder physische Bestrafung.

11. Auszeitbestrafung (time out): Die unerwünschte Verhaltensweise wird dadurch abgeschwächt, daß die Versuchsperson von der verstärkungsreichen Umgebung isoliert wird, z.B. fünfminütige Isolation in einem kleinen leeren Zimmer; Wegräumen der Speise für 15 Sekunden.

12. Überkorrektur (overcorrection): Die unerwünschte Verhaltensweise wird durch massive Übung der alternativen erwünschten Verhaltensweise abgebaut. Z.B. wurde bei einem Fall von Exhibitionismus die Patientin aufgefordert, zusätzliche Kleidung zu tragen und 30 Minuten lang den anderen Patientinnen beim Ankleiden zu helfen.

13. Unterbrechen (interruption): Während des Auftretens einer unerwünschten Verhaltensweise wird verbal oder physikalisch interveniert, um sie zu unterbinden. Z.B. als der Patient bei einem Fall von Epilepsie (seizure) starrte (ein Vor-Epilepsie-Verhalten) schrie der Versuchsleiter laut und scharf "NEIN", ergriff den Patienten mit beiden Händen bei den Schultern und schüttelte ihn kräftig; Bei einem Selbststimulationsverhalten rief der Versuchsleiter scharf "NEIN" und hielt den Körperteil des Kindes fest, um das Verhalten einzustellen.

14. Maßregel (rule): z.B. Klassenregeln, die von allen Schülern eingehaltet werden müssen.

15. Instruktion (Anweisung): Die Vp wird angewiesen, was sie tun soll. Z.B. wurde bei einem Fall von Homosexualität der Patient instruiert, sich nur durch die Diapositive von Frauen erregen zu lassen und die Erregung durch Bilder von Männern zu hemmen.

16. Prompting (prompting; priming): Hinweis, was die Vp tun soll. Z.B. wurde dem Lehrer ein Wink gegeben, sofort soziale Verstärkung zu geben; Verbale oder physikalische

Lernhilfe bei Imitationsverhalten gehört auch zu dieser Kategorie.

17. Imitation: Der Vp wird gezeigt, wie sie tun soll, und dann ahmt sie es nach.

18. Trainingsgeräte (apparatus): Dies sind Geräte, die das Training eines erwünschten Verhaltens unterstützen. Z.B. wurde ein rhythmisches Gerät zur Behandlung von Stottern angewendet; Gerät für Sauberkeitstraining (Patienten, die die Toilette nicht benutzten, wurden mit Hilfe des Gerätes trainiert, dies zu tun).

19. Systematische Desensibilisierung: Dies ist eine psychotherapeutische Technik, die durch Gegenkonditionierung die Angst oder Phobie reduziert. Normalerweise läuft der therapeutische Vorgang in folgender Weise ab: Zunächst wird ein Interview oder Test durchgeführt, um die Angsthierarchie aufzustellen; dann werden unter totaler Entspannung die die unangenehmen Reaktionen auslösenden Reize sukzessiv von dem schwächsten bis zum stärksten eingeführt, bis der Patient keine Angst mehr vor dem die unangenehme Reaktion auslösenden Zielreiz hat.

20. Beschäftigungstherapie: Dem Patienten wird durch anregende Beschäftigung oder tägliche Aktivität, die nur eine Minimalanforderung oder Routine enthält, geholfen.

21. Orgastische Rekonditionierung: Zur Behandlung der Homosexualität wird die heterosexuelle Empfänglichkeit der Patienten durch die Koppelung der heterosexuellen Phantasie oder visuellen Darbietung von Heterosexualität mit dem Orgasmus erhöht.

22. Entspannung: Zur Behandlung der Amnesie oder der Hypertension wird der Patient angewiesen, seine Muskeln zu entspannen und seinen Pulschlag den rhythmischen Schlägen eines Metronomes, das auf 60 Schläge pro Minute gesetzt ist, anzupassen.

23. Gesprächstherapie: Dem Patienten wird durch Gespräche mit dem Therapeuten geholfen.

24. Stimuluskontrolle: Z.B. wird zur Behandlung der Schlaflosigkeit das Bett nur als ein Stimulus für Schlafen betrachtet. Der Patient darf nicht im Bett essen, fern-

sehen oder lesen.

25. Zur Behandlung von Bettnässen wird der Patient zur Blasenentleerung geweckt. Die Häufigkeit des Weckens wird von dreimal bis zu einmal pro Nacht allmählich reduziert.

26. Negative Übung (negative practice): Zur Behandlung des krampfhaften Zuckens(tics) wird der Patient angewiesen, willkürlich und massiv das Zucken wiederholt zu üben.

27. Sonstige: Die unabhängigen Variablen, die den obigen Kategorien nicht zugeordnet werden können, gehören zu dieser Kategorie. Bei einem großen Teil der unabhängigen Variablen in dieser Kategorie handelt es sich um den Vergleich von zwei Behandlungen, d.h. zwei Interventionsstrategien wurden bei einer experimentellen Einzelfalluntersuchung (besonders dem ABAB-Design) verglichen. In diesem Fall war die A-Phase keine Baselinephase, sondern die Treatment-I-Phase und die B-Phase die Treatment-II-Phase, usw. Bei einer solchen Untersuchung muß man von vornherein sicherstellen, daß die unabhängigen Variablen keinen Transfereffekt (Effektverzögerung) haben. In einem solchen Fall interessiert nicht die Frage, ob eine Behandlung einen Effekt hat, sondern ob Behandlung A effektiver als Behandlung B ist, z.B. Selbstkontrolle vs. äußere Kontrolle; verbale Kontingenzt vs. nichtverbale Kontingenzt; Gruppenkontingenzt vs. individuelle Kontingenzt; Timing vs. ohne Timing bei Problemlösung; Anwesenheit vs. Abwesenheit des Versuchsleiters; mit zuvor gegebenen Fragen vs. ohne zuvor gegebene Fragen bei dem Lernen; die Einflüsse des Stressses vs. die der Entspannung auf die Fingertemperaturen, die Hautleitfähigkeit und die Häufigkeit des Pulsschlags; usf.

Tabelle 4-16 und 4-17 geben einen Überblick über die verschiedenen Arten von unabhängigen Variablen aus den Untersuchungen mit Einzelfall-Experimentaldesign. Für die Analyse wurden die unabhängigen Variablen eines Treatmentpakets im einzelnen berechnet. Z.B. bei einem

Tabelle 4-16: Auftretenshäufigkeiten der verschiedenen Kategorien von unabhängigen Variablen bei JABA ¹⁾

Jahrgänge	68	69	70	71	72	73	74	75	76	Σ	%
1.Primäre Verstärkung	4	1	5	2	1	2	3	1	0	19	6.1
2.Soziale Verstärkung	11	5	3	3	5	12	3	3	3	48	15.3
3.Tokensystem	3	2	5	4	13	11	8	2	8	56	17.9
4.Aktivitätskonsequenz	0	5	4	2	2	5	0	0	1	21	6.7
5.Feedback	1	1	1	1	2	2	4	2	2	16	5.1
6.Selbstkontrolle	0	0	0	0	0	1	1	1	0	3	1.0
7.Training	0	1	3	2	0	0	9	9	11	35	11.2
8.Verstärkerentzug	0	2	0	1	5	1	3	1	1	14	4.5
9.Ignorieren	1	1	0	0	1	0	0	0	0	3	1.0
10.Aversive Reize	0	1	0	1	0	0	2	2	0	6	1.9
11.Auszeitbestrafung	1	4	1	2	3	4	2	0	2	19	6.1
12.Überkorrektur	0	0	0	0	0	1	1	0	0	2	0.6
13.Unterbrechen	0	0	0	0	0	0	1	1	0	2	0.6
14.Maßregel	1	1	0	0	0	0	1	0	0	3	1.0
15.Instruktion	0	1	1	1	1	2	0	0	2	8	2.6
16.Anregung	0	1	1	0	1	0	1	1	5	10	3.2
17.Imitation	0	0	0	0	0	1	1	1	0	3	1.0
18.Trainingsgeräte	1	0	0	1	0	0	0	0	0	2	0.6
27.Sonstige	4	3	3	2	3	9	8	3	8	43	13.7

1) Anhang 5 gibt an, in welchen Artikeln die verschiedenen Kategorien der unabhängigen Variablen untersucht wurden.

Tabelle 4-17: Auftretenshäufigkeiten der verschiedenen Kategorien von unabhängigen Variablen bei den drei Zeitschriften (JBTEP, BT und BRT).¹⁾

Zeitschriften	<u>JBTEP</u>	<u>BT</u>	<u>BRT</u>	Σ	%
1. Primäre Verstärkung	4	4	0	8	5.4
2. Soziale Verstärkung	4	10	0	14	9.4
3. Tokensystem	13	13	5	31	20.8
4. Aktivitätskonsequenz	0	0	1	1	0.7
5. Feedback	3	2	1	6	4.0
6. Selbstkontrolle	5	2	0	7	4.7
7. Training	4	7	2	13	8.7
8. Verstärkerentzug	2	2	0	4	2.7
9. Ignorieren	1	2	0	3	2.0
10. Darbietung aversiver Reize	7	1	2	10	6.7
11. Auszeitbestrafung	9	6	3	18	12.1
12. Überkorrektur	0	1	2	3	2.0
13. Unterbrechen	1	0	1	2	1.3
15. Instruktion	2	1	2	5	3.3
17. Imitation	1	1	0	2	1.3
18. Trainingsgeräte	0	0	1	1	0.7
19. Desensibilisierung	3	1	0	4	2.7
20. Beschäftigungstherapie	0	1	0	1	0.7
21. Orgastische Rekonditionierung	0	1	0	1	0.7
22. Entspannung	1	1	0	2	1.3
23. Gesprächstherapie	1	0	0	1	0.7
24. Stimuluskontrolle	1	0	0	1	0.7
25. Wecken	1	0	0	1	0.7
26. Negative Übung	1	1	1	3	2.0
27. Sonstige	2	5	0	7	4.7

1) Anhang 5 gibt an, in welchen Artikeln die verschiedenen Kategorien der unabhängigen Variablen untersucht wurden.

B-BC-B-BC-Design (wobei B=Feedback, C=soziale Verstärkung) wird einmal "Feedback" und einmal "soziale Verstärkung" aufgezeichnet. Unabhängige Variablen, die nicht vollständig untersucht wurden, werden nicht aufgezeichnet. Z.B. bei einem A-B-C-D-C-D-Design werden die unabhängigen Variablen A und B nicht registriert, weil ihre Effekte nicht experimentell überprüft wurden. Beide Tabellen haben einen ähnlichen Trend, d.h. einige unabhängige Variablen tauchen häufiger als die anderen auf. Tokensystem, soziale Verstärkung, Training, Auszeit, bestrafung und primäre Verstärkung haben beispielsweise mehr Auftretenshäufigkeiten. Eine Korrelation zwischen beiden Häufigkeitsverteilungen ergibt $r=0.78$, $N=27$, $p < 0.01$. (Die in einer der beiden Tabellen fehlenden unabhängigen Variablen wurden mit der Häufigkeit 0 bezeichnet).

Die beiden Tabellen besagen, daß sich die Mehrheit der unabhängigen Variablen aus den Untersuchungen mit Einzelfall-Experimentaldesign auf Techniken der Verhaltensmodifikation beschränkt.

6) Wie ist die experimentelle Einzelfalluntersuchung hinsichtlich der abhängigen Variablen verallgemeinert worden?

In der vorliegenden Untersuchung werden die abhängigen Variablen der ausgewählten Artikel in fünf Kategorien klassifiziert:

- A. soziale Verhaltensweisen
- B. Schulleistungsverhalten (academic behavior)
- C. Selbstständigkeitsverhalten (self-care behaviors)
- D. psychisch abweichende Verhaltensweisen
- E. Sonstige

Die zu diesen Kategorien gehörenden Untergruppen werden im folgenden aufgeführt:

A. soziale Verhaltensweisen

1. Aufmerksamkeitsverhalten: unter Aufmerksamkeitsverhalten versteht man das Verharren des Lernenden beim erwünschten Lerngegenstand, z.B. Aufmerksamkeit lenken auf den laufenden Unterricht (beschäftigen mit Schulaufgaben, dem Lehrer zuhören, zweckmäßige Frage stellen, usw.).

2. Pünktlichkeit: ohne Verspätung zur Arbeit, Versammlung oder zum Unterricht kommen.
3. Sauberkeit und Ordnung: Zustand der Sauberkeit und Ordnung halten, z.B. Zimmerputzen, Bettmachen, Abfälle beseitigen (auch in der Öffentlichkeit), Zähne putzen, gepflegtes Aussehen.
4. Pro-soziale Verhaltensweisen: Dazu gehören positive Sozialkontakte (Interaktion mit anderen Kameraden, Interrasseninteraktion, usw.), Kooperation und positive sozial-emotionale Verhaltensweisen (wie Lächeln, positive physikalische Kontakte, verbale Grüße usw.).
5. Partizipation: Teilnahme an einer Aktivität (z.B. Spielen), Veranstaltung oder am Unterricht.
6. Aufgabenausführung: Ausführen der zugewiesenen Aufgaben. Hierunter werden nur die vollendeten Aufgaben gezählt, die Richtigkeit oder Genauigkeit der Arbeit wird nicht berücksichtigt. Durch diese operationale Definition unterscheidet sich diese Subkategorie von der Subkategorie "Schulaufgabe".
7. Soziale Fertigkeiten (social skills): die für die Kommunikation mit anderen notwendigen Fertigkeiten. Zum Training bei den sich isolierenden und zurückziehenden Patienten wurden die sozialen Fertigkeiten in drei Komponenten gegliedert: (1) Blickkontakt mit dem Gesprächspartner, (2) initiative Befragung (initiative questioning), (3) Feedback geben, d.h. soziale Verstärkung geben für das Reden des Gesprächspartners). Andere soziale Fertigkeiten sind: die Fertigkeit, vor einem Publikum in der Öffentlichkeit zu sprechen (public speaking) (Dies bedarf beispielsweise des Blickkontakts mit den Zuhörern, der Verwendung von Gesten und der lebendigen Intonation); die Fertigkeit, sich mit unvernünftigen Ansprüchen anderer assertorisch auseinanderzusetzen; die Fertigkeit, mit dem Ehepartner positiv zu interagieren (z.B. Herstellung des Blickkontakts und Lächelns während der Unterredung).
8. Techniken der Verhaltensmodifikation: Es handelt sich um das Training zur Anwendung von Techniken der Verhaltensmodifikation bei Lehrern, Eltern oder Peers

(z.B. rechtzeitig Verstärkung ausüben). Das Training der Beobachtungstechniken gehört auch zu dieser Subkategorie.

9. Störverhalten (disruptive behavior): Dieses Verhalten kennzeichnet alle mit Lernverhalten inkompatiblen Verhaltensweisen, z.B. hyperaktives Verhalten; Unaufmerksamkeit; mit Nachbarn schwätzen; absichtlich Lärm machen; Lehrer oder Mitschüler beim Lesen, Sprechen unterbrechen; usw.
10. Ungehorsamkeit (noncompliance): liegt vor, wenn ein Kind dem Anspruch oder der Instruktion der Eltern oder des Lehrers nicht gehorcht.
11. Aggressives Verhalten: verbale oder tätliche Angriffe gegen Materialien oder Personen. Zu dieser Subkategorie gehört auch Wutausbruch (laut schreien, auf den Boden stampfen usw.).

B. Schulleistungsverhalten

12. Sprachfähigkeit: Das Erlernen der Sprache und die Behandlung von Sprachstörungen. Das Erlernen der Sprache wird als der Aufbau und die Verstärkung des angemessenen verbalen Ausdruckes aufgefaßt, z.B. das Erlernen der richtigen Anwendung von Substantiven, Adjektiven, Präpositionen, Verben, Singular und Plural, einfachen und komplexen Sätzen usw. Zu der Behandlung von Sprachstörungen, die durch psychische Zustände oder durch Defekte der Sprechorgane verursacht werden, gehört die Behandlung von Stötern, Stammelnen, Aphonie, und Nasalresonanz.
13. Schulaufgabe: Bearbeitung von Aufgaben im Unterricht, wobei die Richtigkeit oder Genauigkeit der Arbeit bewertet wird.
14. Kreatives Verhalten: Es kennzeichnet die Fähigkeit, neue Denk- und Verhaltensmuster zu finden. Die Kreativität wird mit einem bestimmten Maßstab gemessen, z.B. die Anzahl der verschiedenen Ausdrücke beim Aufsatz oder die Anzahl der verschiedenen Formen bei der Blockkonstruktion.
15. Vorschulisches Verhalten(vor-academic behaviors): Vorschulisches Verhalten umfaßt Imitationen und

die Befolgung von Instruktionen (z.B. zeige mir deinen Mund; sage laut "A"; usw.). Da diese beiden Verhaltensweisen Voraussetzung für Lernen sind, werden sie als vorschulische Verhaltensweisen bezeichnet. Wenn ein intellektuell behindertes Kind der Instruktion nicht gehorcht, bedeutet das nicht, daß es nicht gehorsam ist, sondern, daß es die Instruktion nicht versteht.

C. Selbstständigkeitsverhalten (self-care behavior)

16. Ausscheidungsverhalten: die Fähigkeit, selbständig und richtig die Toilette beim Urinieren und Defäkieren zu benutzen. Das Problemverhalten "Bettnässen" gehört auch zu dieser Subkategorie.

17. Eßverhalten: Aufbau angemessenen Eßverhaltens (z.B. die richtige Benutzung des Eßbestecks) und Abbau unangemessenen Eßverhaltens (z.B. mit den Fingern oder unmittelbar mit dem Mund essen; verweigern zu essen).

18. Motorik: die Fähigkeit, selbständig zu stehen, gehen usw. (bes. bei Behinderten).

D. psychisch abweichende Verhaltensweisen

19. Selbstdestruktives Verhalten: z.B. wenn ein psychisch gestörtes Kind seinem eigenen Körper Schaden zufügt.

20. Sexualstörungen: Homosexualität, Exhibitionismus, Impotenz, sexuelle Frigidität, weibliches Verhalten bei Jungen (weibliche Rolle spielen, nur mit Puppen spielen, weibliche Gestik haben, usw.), und übermäßiges Sexualverhalten (z.B. ein Patient hatte einmal die Untreue seiner Frau entdeckt. Seither verließ er seine Dienststelle sehr oft, um zu kontrollieren, ob seine Frau wieder untreu war. Sein sexueller Verkehr mit seiner Frau, den er häufig mit Gewalt erzwang, nahm übermäßig zu).

21. Übermäßige Phantasie und bizarre Sprechweise: verbale Reaktionen, die von der externen realen Situation unabhängig sind, z.B. täuschende Sprache (beharrlich wiederholte und unverständliche Sprechweise) und Halluzination.

22. Phobien: abnorme Objekt- oder Situationsangst, z.B. Schulangst, Klaustrophobie, Messerangst, Angst

- vor Menstruation, Angst vor zahnärztlicher Behandlung, Angst vor Alleinsein, usw.
23. Fallsucht (seizure): Bei dieser Krankheit blickt der Patient zunächst starr auf eine flache Oberfläche, dann wird sein Körper steif, nach dem Krampf fällt er zu Boden.
24. Selbst-Stimulation: rhythmische Bewegungen eines Körperteils (z.B. Hand, Fuß, Kopf oder Gesäß). Dau-menlutschen gehört auch zu dieser Subkategorie.
25. Sucht: zwanghafter Drang nach Lustgewinn, z.B. Alkoholismus, Rauchen, Einatmen von Farbstoffgeruch.
26. Amnesie: Unter Amnesie versteht man den Erinnerungsverlust für alle oder einen Teil der Ereignisse eines Zeitabschnittes. Z.B. bei einer Patientin, die jedesmal, wenn sie getadelt oder kritisiert wurde, sich niederlegte und ihre Augen, Mund und Fäuste schloß und unfähig war, sich an irgendein Ereignis in ihrer Umgebung zu erinnern.
27. Depressives Verhalten: Dieses Verhalten äußert sich in folgenden Symptomen: Schlaflosigkeit, Selbstmordvorstellung (suicidal ideation), sich Zurückziehen, Weinperioden, Nervosität.
28. Scavenging behavior: Es umfaßt pica (essen nicht-nahrhafter Stoffe) und coprophagy (essen von Kot).
29. Zucken (tics): z.B. chronisches, zwanghaftes Zucken mit dem Hals oder ganzen Körper.
30. Zwangsverhalten: unsinnige und unangemessene aber willentlich schwer beeinflussbare Wiederholung einer Handlung, z.B. Waschzwang, übermäßiges Naseputzen, wiederholtes Nachsehen, ob der Gashahn abgedreht, das Zimmer abgeschlossen ist, usw., ERzwang (zwangsmäßiges Essen von Kuchen, Bonbons usw.).
31. Hypertension: Der Zusammenhang zwischen Hypertension und dem Blutdruck, Herzschlag, der Handtemperatur Hautleitfähigkeit wurde untersucht. Man versucht durch Konditionierung den Blutdruck, Herzschlag, die Fingertemperatur, die Hautleitfähigkeit zu normalisieren und damit Hypertension zu verringern. Ray-nauds Krankheit (kalte Hände und Füße) und Kopf-

schmerzen hängen auch mit Hypertension zusammen.

32. Sonstige: Die abhängigen Variablen, die den oben erwähnten Kategorien nicht zugeordnet werden können, werden in diese Kategorie eingeschlossen, z.B. asthmatische Reaktionen im Schlaf, Zähneknirschen im Schlaf, gewohnheitsmäßiges Erbrechen nach der Mahlzeit, unzüchtige Redensarten, Anpassung an normales Leben (z.B. der behinderte Patient wurde trainiert, sich im Straßenverkehr richtig zu verhalten); Verminderung des Fehlbetrags (case shortage) bei einem Kaufhaus; Erhöhung der Anzahl der Busgäste; Verringerung der Anzahl der gemeldeten Einbrüche in einem Wohnviertel; Verbesserung der Ausstattung einer Säuglingsfürsorgeanstalt; Beschleunigung der Tokenentwertung; Förderung der Vorschläge der Patienten; Training zur richtigen Benutzung eines therapeutischen Gerätes; Erhöhung des Umsatzes; Förderung des Konsums von Mittel-Schicht-Essen bei Kindern im ländlichen Kindergarten; Zeitmanagement; und Behandlung von Fett-leibigkeit.

Die Auftretenshäufigkeiten der abhängigen Variablen in den Untersuchungen mit Einzelfall-Experimentaldesign sind in Tabelle 4-18 und 4-19 zusammengestellt. Betrachtet man die beiden Tabellen, so kann man feststellen, daß ungefähr die Hälfte der untersuchten abhängigen Variablen zur Kategorie "soziales Verhalten" gehört (in JABA 60.6% und in den anderen drei Zeitschriften 48.3%). In JABA kommt das Schulleistungsverhalten an zweiter Stelle (26.2%), in den anderen drei Zeitschriften steht an zweiter Stelle die Kategorie "psychisch abweichende Verhaltensweisen" (34.6). Die der Kategorie "Sonstige" zugeordneten abhängigen Variablen machen nur einen geringen Prozentsatz aus (4.3% in JABA und 3.6% in den anderen drei Zeitschriften).

Tabelle 4-18: Auftretenshäufigkeiten der verschiedenen Kategorien von abhängigen Variablen in JABA

Jahrgänge	68	69	70	71	72	73	74	75	76	Σ	%
1.Aufmerksamkeit	3	1	5	2	3	6	5	5	1	31	11.0
2.Pünktlichkeit	0	0	1	1	0	1	0	0	1	5	1.8
3.Sauberkeit und Ordnung	1	0	0	1	0	2	1	2	4	11	3.9
4.Pro-soziales Verhalten	4	1	3	1	2	2	2	0	4	19	6.7
5.Partizipation	0	0	3	0	2	1	3	0	2	11	3.9
6.Aufgabenausführung	0	3	1	1	2	0	3	1	4	15	5.3
7.Soziale Fertigkeiten	0	0	0	0	0	0	0	1	4	5	1.8
8.Techniken zur Verhaltensmodifikation	2	0	2	0	2	2	3	4	2	17	6.0
9.Störverhalten	3	5	3	3	5	7	5	3	3	37	13.1
10.Ungehorsamkeit	0	2	0	0	0	2	1	1	1	7	2.5
11.Agressives Verhalten	2	1	1	1	2	2	1	0	3	13	4.6
<u>1-11. Sozialverhalten</u>											(60.6%)
12.Sprachfähigkeit	6	1	4	1	2	5	6	2	1	28	9.9
13.Schulaufgabe	1	0	4	2	4	3	7	4	5	30	10.6
14.Kreatives Verhalten	0	0	0	0	1	1	0	0	1	3	1.1
15.Vorschulisches Verhalten	1	1	3	3	1	2	2	0	0	13	4.6
<u>12-15. Schulleistungsverhalten</u>											(26.2%)
16.Ausscheidung	0	0	0	2	0	0	0	0	0	2	0.7
17.EBverhalten	0	1	1	0	1	0	0	0	0	3	1.1
18.Motorik	0	0	0	1	1	0	0	1	0	3	1.1
<u>16-18. Selbstständigkeitsverhalten</u>											(2.9%)
19.Selbstdestruktion	0	1	0	1	0	0	0	1	0	3	1.1
20.Sexualstörungen	0	0	0	0	0	1	1	0	0	2	0.7
21.Übermäßige Phantasie und bizarre Sprechweise	0	0	0	0	1	1	0	0	0	2	0.7
22.Phobien	1	0	0	0	0	0	0	0	0	1	0.4
23.Fallsucht	0	0	0	0	0	0	0	1	0	1	0.4
24.Selbststimulation	0	0	0	1	0	2	3	0	0	6	2.1
25.Sucht	0	0	0	0	0	0	0	1	0	1	0.4
27.Depressives Verhalten	0	0	0	0	1	0	0	0	0	1	0.4
<u>19-27. Psychisch abweichende Verhaltensweisen</u>											(6.2%)
3 . Sonstige	0	1	0	0	0	1	5	1	4	12	(4.3%)

(Anhang 6 gibt die Artikel an, die diese Variablen untersuchten)

Tabelle 4-19: Auftretenshäufigkeit der verschiedenen Kategorien von abhängigen Variablen in JBTEP, BT und BRT

Zeitschriften	<u>JBTEP</u>	<u>BT</u>	<u>BRT</u>	Σ	%
1.Aufmerksamkeitsverhalten	1	6	0	7	5.0
3.Sauberkeit und Ordnung	4	2	2	8	5.8
4.Pro-soziales Verhalten	3	5	0	8	5.8
5.Partizipation	3	4	1	8	5.8
6.Aufgabenausführen	1	0	0	1	0.7
7.Soziale Fertigkeiten	2	2	1	5	3.6
8.Techniken zur Verhaltensmodifikation	1	1	0	2	1.4
9.Störverhalten	7	9	0	16	11.5
10.Ungehorsamkeit	1	3	0	4	2.9
11.Aggressives Verhalten	5	3	0	8	5.8
<u>1-11. Soziales Verhalten</u>					(48.3%)
12.Sprachfähigkeit	0	3	0	3	2.2
13.Schulaufgabe	2	3	1	6	4.3
15.Vorschulisches Verhalten	2	0	0	2	1.4
<u>12-15. Schulleistungsverhalten</u>					(7.8%)
16.Ausscheidungsverhalten	2	1	1	4	2.9
17.Erverhalten	1	1	0	2	1.4
18.Motorik	0	2	0	2	1.4
<u>16-18. Selbständigkeitsverhalten</u>					(5.7%)
19.Selbstdestruktives Verhalten	1	1	0	2	1.4
20.Sexualstörungen	3	2	4	9	6.5
21.Übermäßige Phantasie und bizarre Sprechweise	6	2	0	8	5.8
22.Phobien	0	1	0	1	0.7
23.Fallsucht	0	1	0	1	0.7
24.Selbststimulation	2	0	0	2	1.4
25.Sucht	3	1	4	8	5.8
26.Amnesie	2	0	0	2	1.4
27.Depressives Verhalten	2	1	0	3	2.2
28.Scavenging behavior	0	0	1	1	0.7
29.Zucken	2	0	1	3	2.2
30.Zwangsverhalten	2	1	0	3	2.2
31.Hypertension	2	2	1	5	3.6
<u>19-31. Psychisch abweichende Verhaltensweisen</u>					(34.6%)
32.Sonstige	1	2	2	5	(3.6%)

(Anhang 6 gibt die Artikel an, die diese Variablen untersuchten)

4.4 Zusammenfassung der Ergebnisse der Untersuchung II

Insgesamt wurden 367 Artikel aus den vier Zeitschriften zur Verhaltensmodifikation ausgewählt. Obwohl diese vier Zeitschriften in gleicher Weise als Untersuchungen zur Verhaltensmodifikation repräsentieren (Kazdin 1975), orientiert sich das JABA mehr an der schulischen Situation, während die anderen drei Zeitschriften sich mehr mit der klinischen Forschung beschäftigen (s. Tabelle 4-15). Deshalb sind in JABA mehr nichtbehinderte als behinderte Vpn in die Untersuchungen einbezogen, während bei den anderen drei Zeitschriften das Verhältnis umgekehrt ist (s. Tabelle 4-13). Im Hinblick auf die abhängigen Variablen, außer der Kategorie "soziales Verhalten", sind in JABA mehr Artikel, die sich mit Schulleistungsverhalten auseinandersetzen, während in den anderen drei Zeitschriften mehr psychisch abweichende Verhaltensweisen untersucht werden (vgl. Tabelle 4-18 und 4-19).

Obwohl das JABA und die anderen drei Zeitschriften unterschiedliche Orientierungen bezüglich der Untersuchungsvariablen haben, zeigen sie eine hohe Übereinstimmung der Daten hinsichtlich der in der vorliegenden Arbeit untersuchten inneren Validität der experimentellen Einzelfalluntersuchung. In allen vier Zeitschriften haben ungefähr ein Drittel der Phasenpaare das Kriterium von Nicht-Überschneiden erreicht (s. Tabelle 4-6). Dies besagt, daß das Kriterium von Nicht-Überschneiden allein für die Überprüfung der Phasendifferenz nicht ausreichend ist. Man muß andere Prüfverfahren dafür anwenden. Im Hinblick auf die Phasenlänge überwiegen in den vier Zeitschriften kleine Stichproben, d.h. die meisten Autoren erhoben in einer Phase weniger als 30 Beobachtungspunkte (s. Abb. 4-3). Wenn man ein statistisches Prüfverfahren für die Phasendifferenz sucht, muß man dieses Phänomen berücksichtigen. Bei der Berechnung der Beobachtungsübereinstimmung ergibt sich überraschend ein durchschnittlicher Prozentsatz von 93.6%, wobei die Schwankung gering ist ($\sigma=0.85\%$) (s. Abb. 4-4). Dieser Prozentsatz kann als ein Maßstab für das Training der Beobachter angesehen werden.

4.5 Diskussion der Anwendbarkeit der verschiedenen statistischen Prüfverfahren

In diesem Abschnitt werden die in Tabelle 4-7 aufgeführten Prüfverfahren für die Phasendifferenz der experimentellen Einzelfalluntersuchung hinsichtlich ihrer Anwendbarkeit erörtert.

Die Anwendbarkeit eines Prüfverfahrens für die Phasendifferenz der experimentellen Einzelfalluntersuchung hängt davon ab, ob die Eigenschaften der zu überprüfenden Erhebungsdaten den Voraussetzungen des Prüfverfahrens entsprechen oder nicht. Die statistischen Prüfverfahren werden in zwei Gruppen klassifiziert: parametrische und nicht-parametrische statistische Prüfverfahren. Ein parametrisches Prüfverfahren stellt bestimmte Bedingungen an die Parameter der Population, aus der die Beobachtungsstichprobe gezogen wurde, während ein nicht-parametrisches Prüfverfahren keine derartigen Bedingungen stellt (vgl. Siegel, 1976, S. 31). Der t-Test und der F-Test gehören zu den parametrischen Prüfverfahren, die anderen Prüfverfahren sind nicht-parametrisch.

Die Voraussetzungen für t-Test und F-Test lauten:

- 1) Unabhängigkeit der Beobachtungen, d.h. das Auftreten einer Stichprobe aus der Population darf die Auftretens-chancen anderer Stichproben nicht beeinflussen.
- 2) Normalverteiltheit der Population, aus der die Beobachtungen stammen.
- 3) Homogenität der Varianz der Populationen, d.h. die Populationen müssen dieselbe Varianz haben.
- 4) Die Beobachtungen müssen mindestens intervallskaliert sein.
- 5) Ein F-Test verlangt außerdem noch die Additivität der Effekte, d.h. daß keine Wechselwirkung der Versuchsperson mit dem Behandlungsniveau zum Meßwert beiträgt (vgl. Siegel 1976, S. 20).

Die einzige wesentliche Voraussetzung für ein nicht-parametrisches Prüfverfahren ist die Unabhängigkeit der Beobachtungen (vgl. Siegel 1976, S. 31).

Die Verletzung einer Voraussetzung eines statistischen Prüfverfahrens kann zu einem Fehler vom Typ I oder Typ II beim Überprüfen der Hypothesen führen. Aber eine Anzahl

von empirischen Untersuchungen über die Auswirkungen von Verletzungen dieser Voraussetzungen zeigen, daß der F-Test, wie der t-Test, sowohl gegenüber mangelnder Normalverteilung als auch gegenüber unterschiedlichen Populationsvarianzen ziemlich robust ist (vgl. Lewis 1974, S.48).

Scheffé hat nachgewiesen, daß der Effekt der Verletzung der Normalitätsannahme auf die Inferenz der Varianzen sehr groß, dagegen auf die Inferenz der Mittelwerte sehr gering ist (vgl. Scheffé 1959, S.33'). Scheffés empirische Daten zeigen, daß bei ungleicher Gruppengröße ($n_1 \neq n_2$), z.B. $n_1/n_2 = 2, 3, \dots$, oder k , die Heterogenität der Varianz erheblich die wahre Wahrscheinlichkeit des 95%-Konfidenzintervalls für $u_1 - u_2$ beeinflusst, wobei u_1 und u_2 die Mittelwerte der Populationen sind. Aber wenn die Gruppengröße gleich ist, dann spielt die Verletzung der Varianzhomogenität keine Rolle (von $\sigma_2^2/\sigma_1^2 = 0$ bis $\sigma_2^2/\sigma_1^2 = \infty$ haben alle wahren Wahrscheinlichkeiten im 95%-Konfidenzintervall für $u_1 - u_2$ einen Wert von 0.05, wobei σ_2^2 die Varianz der Gruppe 2. und σ_1^2 der Gruppe 1 ist) (vgl. Scheffé 1959, S.40).

Problematisch ist die Verletzung der Unabhängigkeit der Beobachtungen. Scheffé meint, daß die Abhängigkeit der Beobachtungen mit Hilfe der Zeitreihenkorrelation überprüft werden kann. Die Zeitreihenkorrelation wird als Autokorrelation von Lag 1 verstanden, d.h. die Korrelation zwischen x_i und x_{i+1} , wobei $i = 1, 2, \dots, N-1$, N = die Anzahl der Beobachtungen (vgl. Scheffé 1959, S.333-334). Wenn der Koeffizient der Autokorrelation von Lag 1 signifikant größer als null ist, dann sind die Beobachtungen voneinander abhängig. Scheffé hat mit den empirischen Daten den Effekt der Zeitreihenkorrelation auf die wahre Wahrscheinlichkeit im 95%-Konfidenzintervall für u dargestellt. Seine empirischen Daten zeigen, daß bei einem Zeitreihenkorrelationskoeffizienten (r) gleich null die wahre Wahrscheinlichkeit im 95%-Konfidenzintervall 0.05 ist, aber bei $r < 0$ wird die Wahrscheinlichkeit immer geringer, z.B. bei $r = -0.3$ ist die Wahrscheinlichkeit nur 0.002. Wenn $r > 0$ ist, wird die Wahrscheinlichkeit immer größer, z.B. bei $r = 0.2$ beträgt 0.098, und bei $r = 0.4$ ist sie 0.14 (vgl. Scheffé 1959, S.339). Infolgedessen darf die Voraussetzung der Unabhängigkeit der

Beobachtungen nicht verletzt werden.

Bei dem N=1-Design sind die Daten sowohl innerhalb der Gruppe (innerhalb der Phase) als auch zwischen den Gruppen (zwischen den Phasen) abhängig, weil die Daten für eine und dieselbe Versuchsperson erhoben werden.

Um den Effekt der Zeitreihenabhängigkeit der Daten auf die Übereinstimmung zwischen der visuellen und statistischen Inferenz zu untersuchen, haben Jones et al. (1978) 24 verschiedene Experimente aus dem Journal of Applied Behavior Analysis als Stichprobe gezogen. Zwanzig von diesen 24 Experimenten haben statistisch signifikante Lag-1-Autokorrelationen ($p < 0.05$). Dies verdeutlicht, daß die Annahme der Unabhängigkeit der Beobachtungen bei experimentellen Einzelfalluntersuchungen, wie sie von Gentile et al. (1972) und Shine et al. (1971) postuliert wird, nicht akzeptabel ist. Unter Berücksichtigung dieser Tatsache müssen, mit Ausnahme der Zeitreihenanalyse, die im Teil "N=1" der Tabelle 4-7 auftretenden statistischen Prüfverfahren (t-Test für verbundene bzw. korrelierte Stichproben, t-Test, H-Test, Varianzanalyse nach Gentile et al. (1972), Chi-Quadrat-Test, Mann-Whitney-U-Test und Vorzeichen-Test) zur Überprüfung der Phasendifferenz des N=1-Designs als ungeeignet bezeichnet werden, weil solche statistischen Prüfverfahren zumindest die Unabhängigkeit der Beobachtungen innerhalb der Gruppe verlangen, m.a.W. die Beobachtungen in der Phase müssen unabhängig sein.

Das einzige anwendbare statistische Prüfverfahren für N=1-Design ist die Zeitreihenanalyse. Bei der Zeitreihenanalyse transformiert man die Rohdaten nach dem linearen Regressionsmodell in nicht korrelierte oder unabhängige Daten und dann überprüft man die Phasendifferenz der transformierten Daten (Jones et al. 1977).

Jones et al. (1977) meinen, daß die Zeitreihenanalyse neben der visuellen Beurteilung ein wichtiges ergänzendes Instrument für die Überprüfung der Phasendifferenz der experimentellen Einzelfalluntersuchung ist. Sie haben mit Hilfe der Zeitreihenanalyse experimentelle Daten aus dem Journal of Applied Behavior Analysis noch einmal analysiert. Die von ihnen demonstrierten Beispiele umfassen die Verschie-

denartigkeit der Experimente zur Verhaltensmodifikation: Sie enthalten sowohl Umkehr- als auch Multiple-Baseline-Designs; sowohl N=1- als auch N=1-Gruppe-Design; Designs mit ungleicher Anzahl der Beobachtungen in den verschiedenen Phasen (die Anzahl der Beobachtungen in einer Phase variiert zwischen 3 und 34); sowohl signifikante Niveau- als auch signifikante Trendveränderungen (und diese Veränderungen wurden sowohl bei N=3 als bei N=34 gefunden, wobei N = die Anzahl der Beobachtungen in der Phase).

Beim N=1-Gruppe-Design, bei dem zur statistischen Verarbeitung alle Daten jeder Vp in jeder Phase durch den jeweiligen Mittelwert vertreten werden, können die Daten innerhalb der Gruppe (Phase) als unabhängig betrachtet werden, weil jede Vp ein anderes Verhaltensniveau hat und man das Verhaltensniveau einer Vp nicht aus dem Verhaltensniveau anderer Vpn vorhersagen kann. Aber die Daten zwischen den Gruppen (Phasen) sind abhängig geblieben, weil die Daten der beiden Phasen von derselben Gruppe von Vpn erhoben wurden. Unter Berücksichtigung dieser Eigenschaft der Daten des N=1-Gruppe-Designs kann man überprüfen, ob die im Teil "N=1-Gruppe" in Tabelle 4-7 aufgeführten statistischen Prüfverfahren anwendbar sind oder nicht. Beim N=1-Gruppe-Design müssen t-Test für unabhängige Stichproben, Varianz-Analyse für unabhängige Gruppen, Chi-Quadrat-Test für unabhängige Stichproben, Omega-Quadrat-Schätzung für unabhängige Stichproben, und Mann-Whitney-Test als nicht angebracht zur Überprüfung der Phasendifferenz beurteilt werden, weil solche Prüfverfahren nur für zwischen den Gruppen (Phasen) unabhängige Daten anwendbar sind. D.h. die Daten der Behandlungsbedingungen bzw. Behandlungsphasen sollen aus verschiedenen Gruppen von Vpn stammen, aber das ist beim N=1-Gruppe-Design nicht der Fall (siehe Lewis 1974, S.27-49 für t- und F-Test; Siegel 1976, S.101-108 für Chi-Quadrat-Test; Hays 1969, S.327 für Omega-Quadrat-Schätzung; und Siegel 1976, S.112-123 für Mann-Whitney-U-Test).

Die übrigen in Tabelle 4-7 auftauchenden statistischen Prüfverfahren für N=1-Gruppe-Design sind angemessen. Diese Prüfverfahren sind für solche Daten geeignet, die innerhalb der Gruppe (Phase) unabhängig und zwischen den Gruppe (Phasen)

abhängig sind:

1. t-Test für gepaarte Stichproben

Der t-Test für gepaarte Stichproben (matched samples) und der t-Test für Korrelierte Stichproben (correlated samples) können als äquivalent betrachtet werden. Bei diesen beiden Methoden wird der gleiche Standardfehler verwendet (Garrett 1965, S.226).

$$t = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\sigma_{M_1}^2 + \sigma_{M_2}^2 - 2r_{12}\sigma_{M_1}\sigma_{M_2}}}, \quad \text{wobei}$$

$\bar{X}_1$ = der Mittelwert aller Vpn in der Phase 1

$\bar{X}_2$ = der Mittelwert aller Vpn in der Phase 2

σ_{M_1} = der Standardfehler des Mittelwerts der Phase 1

σ_{M_2} = der Standardfehler des Mittelwerts der Phase 2

r_{12} = der Korrelationskoeffizient zwischen den Daten der Phase 1 und denen der Phase 2

Bei der Berechnung muß man bedenken, daß die Daten einer Vp in einer Phase durch ihren Mittelwert repräsentiert werden und deshalb jede Vp nur eine Beobachtung (einen Mittelwert) in jeder Phase hat.

Ein anderer anwendbarer t-Test für gepaarte Stichproben ist die "Differenz-Methode", die für eine kleine Anzahl von Vpn besonders geeignet ist (Garrett 1965, S.227-228):

$$t = \frac{M_D - E(M_D)}{\text{est. } \sigma_{M_D}}, \quad \text{wobei}$$

$$M_D = \frac{\sum D_i}{N}, \quad D_i = \bar{X}_{i(2)} - \bar{X}_{i(1)},$$

$\bar{X}_{i(2)}$ = der Mittelwert der Verhaltensraten der Vp i in der Phase 2 (Behandlungsphase).

$\bar{X}_{i(1)}$ = der Mittelwert der Verhaltensraten der Vp i in der Phase 1 (Baselinephase)

$i = 1, 2, \dots, N$. N = die Anzahl der Vpn.

$E(M_D) = 0$ Aufgrund der Null-Hypothese soll die Differenz zwischen beiden Phasen null sein.

$$\text{est. } \sigma_{M_D} = \sqrt{\frac{\sum_{i=1}^N (D_i - M_D)^2}{(N - 1) N}}$$

2. Varianzanalyse für Wiederholte Messungen

Bei der Varianzanalyse für wiederholte Messungen ist die Abhängigkeit der Daten zwischen zwei Phasen erlaubt.

Das Problem der Abhängigkeit der Daten innerhalb einer Phase wird dadurch gelöst, daß die Beobachtungen in einer Phase durch einen Mittelwert repräsentiert werden und dadurch die Zeitreihenkorrelation innerhalb der Phase verschwunden ist. Aber der F-Test setzt auch die Additivität der Treatmenteffekte voraus; darum soll im folgenden das Problem der Additivität erörtert werden.

Die Nicht-Additivität der Treatments besagt, daß eine Wechselwirkung der Vp mit dem Behandlungsniveau zum Meßwert beiträgt. Das nicht-additive Modell ist:

$$Y_{ij}(\text{Meßwert}) = U(\text{Populationsmittelwert}) + S_i(\text{individuelle Differenz}) + A_j(\text{Treatmenteffekt}) + (SA)_{ij}(\text{Interaktionseffekt der Vp mit dem Treatmentniveau}) + e_{ij}(\text{Fehler}).$$

Bei dem additiven Modell ist $(SA)_{ij} = 0$. Dieser Unterschied führt zur Differenz des EMQ(Erwartungswert des mittleren Quadrats) beider Modelle:

Varianzquelle	additives Modell		nicht-additives Modell	
	EMQ	F	EMQ	F
S(Vp)	$\sigma_e^2 + a\sigma_S^2$	$\frac{MQ_S}{MQ_{SA}}$	$\sigma_e^2 + a\sigma_S^2$	
A(Treatment)	$\sigma_e^2 + n\theta_A^2$	$\frac{MQ_A}{MQ_{SA}}$	$\sigma_e^2 + \sigma_{SA}^2 + n\theta_A^2$	$\frac{MQ_A}{MQ_{SA}}$
SA(Fehler)	σ_e^2		$\sigma_e^2 + \sigma_{SA}^2$	

(Aus Myers 1972, S. 170 und 174)

Aus der obigen Tabelle ist ersichtlich, daß die Größe des erwarteten F-Werts (expected F ratio) des nicht-additiven Modells kleiner als die des additiven Modells ist, weil

z.B. $\frac{15+6}{3+6} < \frac{15}{3}$, wobei $3=\sigma_e^2$, $15=n\theta_A^2$, und $6=\sigma_{SA}^2$.

Infolgedessen ist die Stärke-Effizienz (efficiency) des nicht-additiven Modells schwächer als die des additiven Modells (Myers 1972, S. 179). Aber die schwächere Stärke-Effizienz kann durch die Vermehrung der Anzahl der Vpn verbessert werden. Deshalb ist das nicht-additives Modell der Varianzanalyse für wiederholte Messungen auch anwendbar, wenn man sich nicht für den Effekt der Versuchspersonen, sondern nur für den Treatmenteffekt interessiert. Das Rechenverfahren des nicht-additiven Modells ist nicht anders als das des additiven Modells.

Ein anderes potentiell Problem bei der Varianzanalyse für wiederholte Messungen ist die Verfälschung des Treatmenteffekts durch den Übungseffekt. Normalerweise verlangt man bei der Varianzanalyse für wiederholte Messungen, daß die Reihenfolge der verschiedenen Treatmentniveaus (Phasen) für jede Vp randomisiert oder ausbalanciert wird, um den Übungseffekt für jedes Treatmentniveau auszugleichen.

Bei der experimentellen Einzelfalluntersuchung ist die Randomisierung der Reihenfolge der Phasen ausgeschlossen, aber der Übungseffekt kann durch Vermehrung der Beobachtungspunkte vermindert und durch die Umkehrung zur Baselinephase überprüft werden (wenn sich die Kurve der zweiten Baselinephase in die Richtung der ersten Baselinephase

annähert, dann bedeutet es keinen Übungseffekt). Den Übungseffekt muß man bei einer Untersuchung über Schulleistungsverhalten oder über das Erlernen einer Geschicklichkeit berücksichtigen, aber bei einer Untersuchung über soziales Verhalten, besonders wenn die Vp nicht weiß, daß sie beobachtet wird, ist der Übungseffekt gering.

3. Korrelation zwischen unabhängigen und abhängigen Variablen

Kazdin und Klock (1973) haben eine Matrix von Korrelationen zwischen dem Aufmerksamkeitsverhalten von jeder der 12 Vpn und 8 Lehrerverhaltensweisen demonstriert. Diese Methode bietet eine Möglichkeit, zu untersuchen, welche Lehrerverhaltensweisen einen signifikanten Zusammenhang mit der abhängigen Variable haben. Aber die Aussagefähigkeit dieser Methode über die Kausalität ist schwach, weil die verschiedenen Verhaltensweisen des Lehrers nicht manipuliert wurden.

4. Wilcoxon-Vorzeichenrang-Test für abhängige Paare, McNemar-Chi-Quadrat-Test und Randomisierungstest für parallele Paare gehören zur nicht-parametrischen Statistik. Diese drei Methoden sind für das N=1-Gruppe-Design angebracht, wenn die Beobachtungen jeder Vp in jeder Phase durch ihren Mittelwert repräsentiert werden. Diese drei Methoden sind verteilungsfrei und ihre Berechnungsverfahren sind einfacher als die der parametrischen Methoden (für die Berechnungsverfahren siehe Siegel 1976, S.72-80 für den Wilcoxon-Vorzeichenrang-Test für abhängige Paare; S.85-89 für den Randomisierungstest für abhängige Paare, und S.60-65 für den McNemar-Chi-Quadrat-Test). Diese drei Methoden haben auch Nachteile: Weil der McNemar-Chi-Quadrat-Test und der Wilcoxon-Vorzeichenrang-Test für abhängige Paare die aus dem Experiment erhobenen Daten nicht voll ausnutzen, beträgt die Stärke-Effizienz der beiden Methoden im Vergleich zum t-Test nur ungefähr 95% (vgl. Siegel 1976, S.65 und S.80). Der Randomisierungstest für abhängige Paare schöpft die in den Daten enthaltenen Informationen voll aus, deshalb hat er eine Stärke-Effizienz von 100% im Vergleich zum t-Test (Siegel 1976, S.89), aber wenn die Anzahl der Vpn größer als 12 ist, ist das Berechnungsver-

fahren dieser Methode sehr umständlich (vgl. Siegel 1976, S. 88).

5. Zufallsblöcke-Varianzanalyse (Randomized Blocks Analysis of Variance)

Bei der Zufallsblöcke-Varianzanalyse (s. Edwards 1971, S. 179-199) werden Blöcke so gebildet, daß innerhalb eines Blocks die Vpn in bezug auf eine für die Untersuchung relevante Variable möglichst homogen sind. Man kann dieses Design als eine Versuchsanordnung für parallelisierte Stichproben betrachten, besonders wenn eine Vp als Block aufgefaßt wird (s. Edward 1971, S.180, Fußnote 1). In diesem Fall ist die Zufallsblöcke-Varianzanalyse eine Versuchsanordnung, deren Stichproben innerhalb der Phase unabhängig und zwischen den Phasen abhängig sind und deshalb für das N=1-Gruppe-Design angemessen ist.

Zusammenfassend ist die Zeitreihenanalyse für die Überprüfung der Phasendifferenz des N=1-Design angemessen, weil sie die abhängigen Beobachtungen in einer Phase nach dem Linearen-Regressionsmodell in unabhängige Daten transformiert, und dadurch die Voraussetzung der Unabhängigkeit der Beobachtungen erfüllt. Für das N=1-Gruppe-Design sind t-Test für korrelierte Stichproben, Varianzanalyse für wiederholte Messungen, Wilcoxon-Vorzeichenrang-Test für abhängige Paare, McNemar-Chi-Quadrat-Test, Randomisierungstest und Zufalls-Blöcke-Varianzanalyse unter der Bedingung anwendbar, daß die Beobachtungen in einer Phase durch ihren Mittelwert repräsentiert werden. Es wird vorgeschlagen, daß solche anwendbare statistischen Prüfverfahren als ein ergänzendes Instrumentarium benutzt werden. Man sollte nicht auf die visuelle Beurteilung verzichten, weil durch sie Trendveränderungen und Niveauverschiebungen schnell, wenn auch grob, erfaßt werden können.

Jones et al. (1977) haben die Anwendungsmöglichkeit der Zeitreihenanalyse für das N=1-Design demonstriert, aber das ausführliche Verfahren haben sie nicht dargestellt. Box & Jenkins meinen, daß für die richtige Modell-Identifikation mindestens 50 Beobachtungen erforderlich sind (Box & Jenkins 1970, S.33). Aber die meisten experimentellen Einzelfalluntersuchungen umfassen viel weniger als 50 Beobachtungs-

punkte. Jones et al. (1977) haben dieses Problem nicht erörtert. Obwohl schon einige Bücher über Zeitreihenanalyse erschienen sind (z.B. Glass et al. 1975; Box & Jenkins 1970; Huber 1978; Petermann 1977b, S.155-192; Petermann 1978, S.94-99; Petermann & Hehl 1979) ist sie für Erziehungswissenschaftler noch schwer anwendbar. Um die Anwendung der experimentellen Einzelfalluntersuchung zu verbreiten, muß das Berechnungsverfahren der Zeitreihenanalyse in Kochbuchform für Erziehungswissenschaftler geschrieben werden, obgleich es Computer-Programme dafür gibt.

5. Diskussion und Schlußfolgerungen: Wie kann die experimentelle Einzelfalluntersuchung für die pädagogische Forschung fruchtbar gemacht werden?

Zu der Fragestellung: "Was die Psychologie der Erziehung anbieten kann?" hat Bijou (1970) die Meinung vertreten, daß die verhaltenstheoretisch orientierten Wissenschaftler folgendes beitragen können:

- (1) eine Reihe von Konzepten (concepts) und Prinzipien, die ausschließlich durch experimentelle Untersuchungen entstanden sind.
- (2) eine Methodologie für die unmittelbare Umsetzung dieser Konzepte und Prinzipien in die Unterrichtspraxis.
- (3) eine Untersuchungsanordnung, die die Verhaltensänderungen des Kindes im Einzelnen untersucht, anstatt sie aus den Gruppenmittelwerten zu erschließen.
- (4) eine Wissenschaftsphilosophie der Verhaltensanalyse, die auf der beobachtbaren Darstellung der Beziehungen zwischen dem individuellen Verhalten und seinen determinierenden Konditionen beruht.

Nach Bijou ist die Philosophie der Wissenschaft sehr wichtig. Sie beeinflusst die Forschungsthemen, die Methode der Datenerhebung, die Darstellungsformen der Daten und die Interpretation der Ergebnisse. Bijou (1970) hat fünf grundlegende Annahmen (der Philosophie der Wissenschaft) für eine Verhaltensanalyse dargestellt:

- (1) Gegenstand der Verhaltensanalyse sind die Interaktionen zwischen dem Verhalten eines integralen Organismus und den Umgebungsereignissen. Diese Interaktionen werden in beobachtbaren, meßbaren und replizierbaren Begriffen analysiert.
- (2) Diese Interaktionen sind gesetzmäßig.
- (3) Der Gegenstand ist kontinuierlich, z.B. äußern sich die Kontinuitäten in den Beziehungen zwischen der normalen und der pathologischen Entwicklung; in den Problemen und Verfahren der Grundlagen- und angewandter Forschung; und in der Analyse des psychologischen Phänomens von Rohdaten bis zur theoretischen Formulierung.
- (4) Komplexe Interaktionen entwickeln sich aus einfachen Interaktionen.

- (5) Eine psychologische Theorie und ihre Technologie sind offene und flexible Systeme. D.h. ein neuer Begriff, ein neues Prinzip, oder eine neue Technik kann irgendwann ~~der~~ bestehenden Liste hinzugefügt werden, wenn die Zuverlässigkeit gezeigt werden kann.

Was Bijou (1970) über das Angebot der Psychologie an die Erziehung schreibt, entspricht der Fragestellung dieses Abschnittes: Wie kann die experimentelle Einzelfalluntersuchung für die pädagogische Forschung fruchtbar gemacht werden? Was man bisher mit der experimentellen Einzelfalluntersuchung gemacht hat, besonders im Bereich des sozialen Verhaltens und Schulleistungsverhaltens, ist schon in der Untersuchung II dieser Arbeit dargestellt worden. Wir müssen einerseits aufgrund der "Wissenschaftsphilosophie der Verhaltensanalyse" mit dieser Versuchsanordnung (der experimentellen Einzelfalluntersuchung) die Interaktionen zwischen dem individuellen Verhalten und seinen Umgebungsereignissen weiter untersuchen, um neue Begriffe, Prinzipien und Technologien zu bilden, und um daraufhin komplexe Verhaltensweisen untersuchen zu können, andererseits müssen wir die Methodologie der experimentellen Einzelfalluntersuchung weiter explorieren, um zu beschleunigen, daß diese Versuchsanordnung bei Erziehungswissenschaftlern akzeptiert wird. Im folgenden werden einige methodologische Probleme der experimentellen Einzelfalluntersuchung hinsichtlich der Untersuchung des sozialen und des Schulleistungsverhaltens erörtert.

5.1. Die Verwendung der experimentellen Einzelfalluntersuchung bei der Erforschung des sozialen Verhaltens

Die meisten Forschungsarbeiten über soziale Verhaltensweisen gründen auf den Begriffen und Prinzipien der Verhaltensanalyse, die Bijou (1970) definiert. Die Begriffe der Verhaltensanalyse beziehen sich nach Bijou nur auf die beobachtbaren Verhaltensereignisse und die Umgebungsbedingungen, und auf die Zusammenhänge zwischen ihnen.

Die Verhaltensereignisse werden in zwei Arten gegliedert: das respondente Verhalten, das durch den vorhergehenden Stimulus bestimmt ist, und das operante Verhalten, das durch

den nachfolgenden Stimulus bestimmt ist. Auch die Umgebungsbedingungen lassen sich in zwei Kategorien einteilen: die Stimuli mit funktionellen Eigenschaften (z.B. ein konditionierter Stimulus in einer respondenten Interaktion und ein diskriminativer Stimulus in einer operanten Interaktion) und die Hintergrundfaktoren (z.B. Sättigung und Deprivation bezüglich der verstärkenden Stimuli).

Die Prinzipien der Verhaltensanalyse stellen die funktionellen Zusammenhänge zwischen den Verhaltens- und Umgebungsereignissen dar, z.B. Erwerb und Löschung des operanten Verhaltens; Stimuluskontrolle des operanten Verhaltens; konditionierte Verstärker; Verstärkungsplan; respondentes Verhalten und respondente Konditionierung; aversive Kontrolle; Emotion und Motivation (vgl. Bijou 1970, S.67).

Die Methodologie der experimentellen Einzelfalluntersuchung für die experimentelle Forschung in natürlichen Situationen haben Bijou et al. (1969) von unsystematischer Beobachtung, über Verhaltenskodifizierung (behavior coding), Reliabilität der Beobachtung, bis Durchführung des Experiments ausführlich beschrieben (siehe auch Bijou et al. 1968). Im folgenden sind einige spezielle methodologische Probleme zu diskutieren:

1. Zeitstichprobenverfahren

Beobachtung und Aufzeichnung des Verhaltens sind Verfahren der Verhaltensanalyse zur Quantifizierung des Verhaltensphänomens. Im allgemeinen gibt es drei Aufzeichnungsmethoden: die Frequenzaufzeichnung, die Intervallaufzeichnung und die momentane Zeitstichprobe (s. Powell et al. 1975 und Repp et al. 1976). Bei der Frequenzaufzeichnung werden die Daten innerhalb einer ganzen Sitzung oder eines großen Teils der Sitzung gesammelt, und die Daten werden als Auftretensfrequenz dieser Sitzung oder Zeiteinheit registriert. Bei der Intervall-Aufzeichnung wird eine Sitzung in mehrere gleiche Perioden unterteilt, der Beobachter beobachtet das Verhalten innerhalb dieses Intervalls. Am Ende jedes Intervalls zeichnet er auf, ob das Verhalten innerhalb eines Intervalls aufgetreten ist oder nicht. Man kann auch für das Aufzeichnen ein Aufzeichenintervall nach dem Beobachtungsintervall einfügen, z.B. nach einem 30-Sekunden-Beobachtungs-

intervall registriert der Beobachter im folgenden 10-Sekunden-Aufzeichenintervall, ob das beobachtete Verhalten im letzten Beobachtungsintervall eingetreten ist oder nicht. Das Auftreten des Verhaltens kann man operational so definieren, daß es während der ganzen Sitzung vorkommt, oder daß es zumindest in einem Teil der Sitzung erscheint. Bei der momentanen Zeitstichprobe ignoriert der Beobachter das Verhalten für eine relativ lange Periode, und dann, nach einem Signal, beobachtet er das Verhalten und zeichnet es auf. Das Auftreten des Signals kann regelmäßig oder unregelmäßig sein.

Repp et al. (1976) haben ein Experiment durchgeführt, um festzustellen, ob die gleichen Daten durch verschiedene Zeitstichprobenmethoden gleich repräsentiert werden. Sie verglichen (a_1) 10"/5"-Intervall-Verfahren (d.h. 10-Sekunden-Beobachten und 5-Sekunden-Aufzeichnen), (a_2) 10"/0"-Intervall-Verfahren (d.h. innerhalb jedes 10-Sekunden-Intervalls muß der Beobachter beobachten und aufzeichnen), (b_1) 5"-Momentane-Zeitstichprobe, (b_2) 10"-Momentane-Zeitstichprobe miteinander unter drei Konditionen der Auftretensrate:

(c_1) hoch (10 Pulse/Min.), (c_2) mittel (1 Puls/Min.), und (c_3) niedrig (0.1 Puls/Min.). Die Ergebnisse sind:

- (1) das 10"/0"- und 10"/5"-Intervall-Verfahren differieren nur sehr gering ($< 3\%$) bei jeder Auftretensrate.
- (2) die Daten, die durch die momentane Zeitstichprobenmethode erhoben wurden, konnten die Population nicht richtig repräsentieren.
- (3) die absolute Differenz zwischen der 10-Sekunden-Momentane-Zeitstichprobe und den Intervallmethoden verringert sich, wenn die Auftretensrate des Verhaltens abnimmt. Wenn das untersuchte Verhalten eine hohe Auftretensrate hat, muß man entweder das Frequenzaufzeichnungsverfahren oder die Intervallmethode mit sehr kleinen Intervallen anwenden.

Powell et al. (1975) haben auch die verschiedenen Zeitstichprobenmethoden verglichen. Die Auftretensrate des zu beobachtenden Verhaltens war bei ihrer Untersuchung viel niedriger als die bei der Untersuchung von Repp et al. (1976). Die Forschungsergebnisse von Powell et al (1975) zeigen, daß die Daten, die durch die momentane Zeitstichprobenmethode

erhoben wurden, die Population gut repräsentieren konnten, wenn die Länge eines Intervalls (innerhalb einer 20-Minuten-Sitzung) nicht länger als 80 Sekunden war; war das Intervall länger als 120 Sekunden, wurden große Differenzen zwischen ihnen beobachtet. Je kürzer das Intervall, desto größer die Übereinstimmung zwischen der Stichprobe und ihrer Population. Dieses Prinzip gilt auch für die Intervall-Aufzeichnungsmethoden. Aus den Ergebnissen der zwei gerade regierten Untersuchungen kann man den Schluß ziehen, daß die Intervall-Aufzeichnungsmethode mit einem Intervall kürzer als 80 Sekunden sehr angemessen für die Untersuchung des sozialen Verhaltens ist. Je kürzer das Intervall ist, desto besser, solange der Beobachter keinen Zeitdruck fühlt.

2. Faktoren, die die Beobachtungsreliabilität und/oder Beobachtungsvalidität beeinträchtigen können

Beobachtungsreliabilität wird operational definiert als der Übereinstimmungsgrad zwischen Beobachtern, und die Beobachtungsvalidität als der Übereinstimmungsgrad zwischen den beobachteten Daten und den Standarddaten, die beispielsweise vom Forscher selbst oder von erfahrenen Beobachtern valide aufgezeichnet werden. Kazdin (1977) erwähnt einige Faktoren, die die Beobachtungsreliabilität und/oder -validität beeinträchtigen können:

- a. Reaktivität auf die Reliabilitätsprüfung (reliability assessment): Wenn der Beobachter weiß, "wann" und "von wem" seine Beobachtungsreliabilität überprüft wird, wird die Reliabilität täuschend erhöht. Außerdem machen Kent et al. (1977) darauf aufmerksam, daß die Reliabilität hochverschoben wird, wenn der Forscher zur Zeit der Reliabilitätsprüfung abwesend ist (Betrügerei). Romanczyk et al. (1973) führen diese Phänomene darauf zurück, daß der Beobachter seine Beobachtungskriterien modifiziert, damit seine Daten mit denen des Kontrolleurs (assessor) übereinstimmen. Kent et al. (1977) schlagen vor, die Reliabilitätsprüfung heimlich und vom Forscher beaufsichtigt vorzunehmen.
- b. Definitionsabweichung beim Beobachter (observer drift): Im Laufe der Untersuchung ist es möglich, daß der Beobachter selbst die ursprüngliche Definition des Verhaltens

verändert. Wenn zwei Beobachter stets zusammenarbeiten und miteinander kommunizieren, können sie auch zusammen von der ursprünglichen Definition des beobachteten Verhaltens abweichen. Um diesem Problem entgegenzuwirken, kann man die Beobachter die Definition des Verhaltens ausführlich aufschreiben und mit Hilfe eines Videorecorders periodisch trainieren lassen.

- c. Die Kompliziertheit der Verhaltenskategorien: Je größer die Anzahl der zu beobachtenden Kategorien, desto niedriger ist die Beobachtungsvalidität und -reliabilität. Die in der Trainingsphase geschätzte Beobachtungsvalidität und -reliabilität können höher sein als die in den aktuellen Untersuchungssitzungen, wenn die Beobachtungsstichproben in den realen Sitzungen weniger vertraut, komplexer und weniger vorhersagbar sind. Deshalb darf die Anzahl der zu beobachtenden Verhaltenskategorien nicht zu groß sein. Gegebenenfalls muß die Trainingszeit verlängert werden.
 - d. Ausfallendes Intervall: Wenn für einen der Beobachter ein Intervall ausfällt, kann es sein, daß alle nachfolgenden Intervalle nicht mehr zusammenpassen (vgl. Hartmann 1977). Z.B. Während Beobachter 1 das 10. Intervall aufzeichnet, zeichnet Beobachter 2 das 11. Intervall auf. Dies wird zu niedriger Beobachterübereinstimmung führen. Dieses Problem kann durch die nicht-kostspielige Anwendung eines Cassettenrecorders vermieden werden, der dem Beobachter die Nummer des aufzuzeichnenden Intervalls signalisiert.
3. Die Schätzungsverfahren der Beobachtungsreliabilität
Hartmann (1977) stellt eine 2x2 Tabelle zur Berechnung der Beobachtungsreliabilität vor:

		Beobachter 2	
		(0)	(1)
Beobachter 1	(1)	A	B
	(0)	C	D

$N = A + B + C + D$

wobei (0) = Nicht-Auftreten der Verhaltenskategorie
(1) = Auftreten der Verhaltenskategorie

Für Schätzverfahren der Beobachtungsreliabilität können mit dieser Tabelle die Berechnungen durchgeführt werden:

(a) Konventionelles Verfahren: Prozentsatz der Beobachtungsübereinstimmung = $\frac{B+C}{N} \cdot 100$

(b) Effektiver Prozentsatz der Beobachtungsübereinstimmung:

(b₁) Übereinstimmung über Auftretensprozentsatz (occurrence agreement) = $\frac{B}{A+B+C} \cdot 100$

(b₂) Übereinstimmung über Nicht-Auftretensprozentsatz (nonoccurrence agreement) = $\frac{C}{A+C+D} \cdot 100$

(c) Phi (ϕ) = $\frac{BC-AD}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$

(d) Kappa (k) = $\frac{p_o - p_c}{1 - p_c}$, wobei $p_o = \frac{B+C}{N}$
 $p_c = \frac{(C+D)(A+C)(A+B)(B+D)}{N^2}$

p_c ist die Proportion der zufälligen Übereinstimmung (chance agreement), p_o ist die Proportion der Übereinstimmung zwischen den Beobachtern, und der k-Koeffizient ist der Übereinstimmungsprozentsatz, nachdem die Proportion der zufälligen Übereinstimmung ausgeschaltet worden ist (vgl. Cohen 1960).

Kratochwill & Wetzel (1977) meinen, daß die Methode "Übereinstimmung über den Auftretensprozentsatz" angewendet werden soll, wenn eine niedrige Auftretensrate des Verhaltens (z.B. niedriger als 20%) vorliegt. Im Gegensatz dazu soll die Methode "Übereinstimmung über den Nicht-Auftretensprozentsatz" eingesetzt werden, wenn das Verhalten eine hohe Auftretensrate (z.B. 30%) hat. Kratochwill & Wetzel (1977) geben zu, daß es einige potentielle Begrenzungen in der Anwendung der beiden Verfahren von "Effektivem Prozentsatz der Beobachtungsübereinstimmung" gibt. Z.B. ist im allgemeinen die Auftretensrate des Verhaltens nicht stetig, sie ändert sich zwischen der Baseline- und Behandlungsphase, wenn ein Behandlungseffekt vorliegt. Außerdem ist die Streitfrage: "Welches Niveau des Übereinstimmungsindex ist akzeptabel?" noch nicht beigelegt.

Cohen (1960) hat eine Formel zur Überprüfung der Sig-

nifikanz des k-Koeffizienten (gegenüber der Null-Hypothese) entwickelt:

$$\text{Kritische Ratio} = \frac{k - 0}{\sqrt{\frac{p_c}{N(1-p_c)}}$$

Der sich ergebende Wert soll mit der Wahrscheinlichkeit der Normalverteilung verglichen werden. Aber Cohen (1960) hält es nicht für sinnvoll, die Signifikanz des Kappa-Koeffizienten zu überprüfen, weil ein signifikanter Übereinstimmungskoeffizient für die Validität der Untersuchung nicht ausreichend ist. Optimal soll der Übereinstimmungskoeffizient das 100%-Niveau erreichen. Hartmann (1977) hält 80% beim konventionellen Verfahren (Prozentsatz der Beobachtungsübereinstimmung) und 0.60 bei den Phi- und Kappa-Koeffizienten für akzeptabel. Aber nach dem Ergebnis der Untersuchung II der vorliegenden Arbeit scheinen solche Niveaus zu niedrig zu sein (siehe Abb. 4-4).

4. Lehrer als Beobachter

Bei den in natürlichen Situationen durchgeführten Untersuchungen werden die Lehrer oft gebeten, gleichzeitig die Rolle von Mediator und Beobachter zu übernehmen. Kubany & Sloggett (1973) haben ein momentanes Zeitstichprobenverfahren mit 4-, 8-, oder 16-minütigen variablen Intervallen zur Beobachtung und Aufzeichnung für den Lehrer, der gleichzeitig als Mediator und Beobachter tätig ist, vorgeschlagen. Ihrer Meinung nach, ist das 4-minütige variable Intervall zu verwenden, wenn die Untersuchung täglich nur eine Stunde dauert. Aber wenn die Beobachtung den ganzen Tag dauert, ist das 8- oder 16-minütige variable Intervall anzuwenden. Der Zeitplan (schedule) wird vorher festgelegt und im Laufe der Beobachtung wird die Zeit durch einem Zeitmesser (timer) signalisiert. In dem Moment des Auftretens des Signals beobachtet der Lehrer das Verhalten der Zielperson, zeichnet es auf und stellt den Zeitmesser auf die nächste Zeitstichprobe ein. Die Neueinstellung des Zeitmessers kann auch von einem Schüler vorgenommen werden. Kubany & Sloggett

(1973) haben mit diesem Variablen-Momentanen-Intervall-Zeitstichprobenverfahren demonstriert, daß der Lehrer drei Verhaltenskategorien beobachten und aufzeichnen kann, ohne daß der Unterricht beeinträchtigt wird. Obwohl das Verfahren, den Lehrer gleichzeitig als Mediator und Beobachter einzusetzen, einige Vorteile mit sich bringen kann, (z.B. es kostet weniger und kann die Reaktivität der Vpn auf die Anwesenheit des fremden Beobachters verringern), hat dieses Verfahren den beträchtlichen Nachteil, daß die Reaktivität des Lehrers (er ändert während des Beobachtungszeitraums sein Verhalten) wiederum das Verhalten der Vpn beeinflusst. Die Untersuchung von Hay et al. (1977) zeigt: (1) Bei den Vpn, die vom Lehrer beobachtet wurden, traten größere Veränderungen in angemessenem Verhalten auf als bei den Kontrollversuchspersonen, die von einem anderen Beobachter beobachtet wurden. (2) Die vom Lehrer beobachteten Vpn bekamen mehr Lernhilfen in der Phase, in der der Lehrer gleichzeitig als Mediator und Beobachter tätig war, als in der Phase, in der er nur als Mediator tätig war (Die Verbalisierungen des Lehrers, die zusätzliche Informationen vermittelten, und die Lenkung der Aufmerksamkeit der Vpn auf die Schularbeit wurden als Lernhilfen aufgezeichnet). Die Häufigkeitsdifferenz zwischen beiden Phasen war signifikant. Dieser Effekt wurde bei den Vpn, die nicht vom Lehrer beobachtet wurden, nicht gefunden.

Forehand (1973) hat mit A-B-Design den Mechanismus dieser Art von Reaktivität untersucht (sein Design ist vor-experimentell, deshalb kann seine Untersuchung nur als hypothesenbildend betrachtet werden), und kam zu dem Schluß, daß der Lehrer beim Identifizieren, Beobachten und Aufzeichnen eines Problemverhaltens vorhergehende und nachfolgende Bedingungen, die das Problemverhalten aufrechterhalten, erkennen konnte. Diese Information (Feedback) war für ihn ausreichend, um bezüglich des Problemverhaltens Maßnahmen zu ergreifen, die vom Forscher nicht erwartet wurden.

Infolgedessen ist es nicht empfehlenswert, den Lehrer gleichzeitig als Mediator und Beobachter einzusetzen. Die Anwesenheit eines fremden Beobachters verursacht wahrscheinlich nur geringe Reaktivität bei den Vpn. Johnson & Bolstad

(1975) haben die An- und Abwesenheit des Beobachters verglichen. Es wurde keine signifikante Differenz in den Verhaltensweisen des Kindes und der Mediatorin (der Mutter) während der An- und Abwesenheit des Beobachters gefunden. In der Untersuchung von Johnson & Bolstad (1975) wußte die Mediatorin, daß sie und ihr Kind beobachtet wurden. Ob die von Johnson & Bolstad gemachte Aussage über die Reaktivität der Vpn auch unter der Bedingung gültig ist, daß die Vpn (sowohl Mediator als auch Zielkind) nicht wissen, daß sie beobachtet werden, bedarf weiterer empirischen Beweise.

5.2. Die Verwendung der experimentellen Einzelfalluntersuchung bei der Erforschung des Schulleistungsverhaltens

Das Schulleistungsverhalten wird operational definiert als der Prozentsatz der richtig gelösten Aufgaben in einem Test. Bei den meisten Leistungstests werden Schulleistungen verlangt.

In der Literatur gibt es weniger Berichte über die Methodologie der experimentellen Einzelfalluntersuchung für Schulleistungsverhalten als für soziales Verhalten. Deshalb ist es notwendig, die in der Untersuchung II gesammelten Daten ausführlich zu analysieren, um die Anwendungsmöglichkeit der experimentellen Einzelfalluntersuchung für das Schulleistungsverhalten zu diskutieren. Zu diesem Zweck werden 36 Artikel, die die Schulaufgaben als abhängige Variable untersuchten (siehe Anhang 3), als repräsentative Stichprobe für Schulleistungsverhalten analysiert.

Die Fragestellungen lauten folgenderweise:

1. Bei welchen Kursen ist die experimentelle Einzelfalluntersuchung angewendet worden?
2. Welche Designs sind für das Schulleistungsverhalten angemessen?
3. Wieviel Beobachtungspunkte umfassen die einzelnen Phasen? Um die Beobachtungsdaten des Schulleistungsverhaltens zu erheben, muß man für jeden Beobachtungspunkt einen Test durchführen. Liegen deshalb weniger Beobachtungen des Schulleistungsverhaltens vor?

Abb. 5-1 zeigt, daß Mathematik und Lesen-und-Schreiben größere Häufigkeiten haben. Mathematik enthält das Lösen

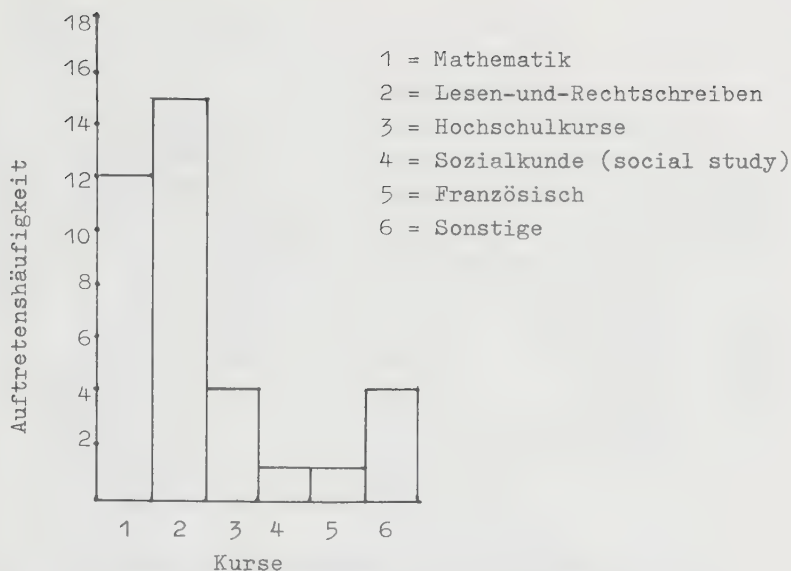


Abb. 5-1: Auftretenshäufigkeit der verschiedenen Kurse, bei denen das Schulleistungsverhalten mit der experimentellen Einzelfalluntersuchung untersucht wurde.

mathematischer und arithmetischer Aufgaben. Lesen-und-Schreiben umfaßt Handschrift, Leseverständnis (comprehension reading), Worterkennung, Geschichten-Schreiben, Biographie-Schreiben, usw. Hochschulkurse sind Psychologie, Statistik, usw. Vier Testaufgaben (Bilder-Malen, follow-the-dots, Test über Fernsehnachrichten und Problem-Lösung) sind nicht in der Schule durchgeführt worden, aber sie haben ähnliche Eigenschaften wie Schulaufgabentests, deshalb werden sie der Kategorie "Sonstige" zugeordnet.

Normalerweise kann man die Materialien (die Lehrstoffe) in mehrere Einheiten aufteilen. Bei jeder Untersuchungssitzung wird eine Einheit gelehrt bzw. gelernt. Aus jeder Einheit werden beispielsweise ca. 10-20 Testaufgaben herausgezogen. Am Ende der Sitzung wird der informelle, kleine Test durchgeführt. Der Prozentsatz der richtig gelösten Aufgaben wird berechnet, und als das Beobachtungsdatum der Sitzung bezeichnet.

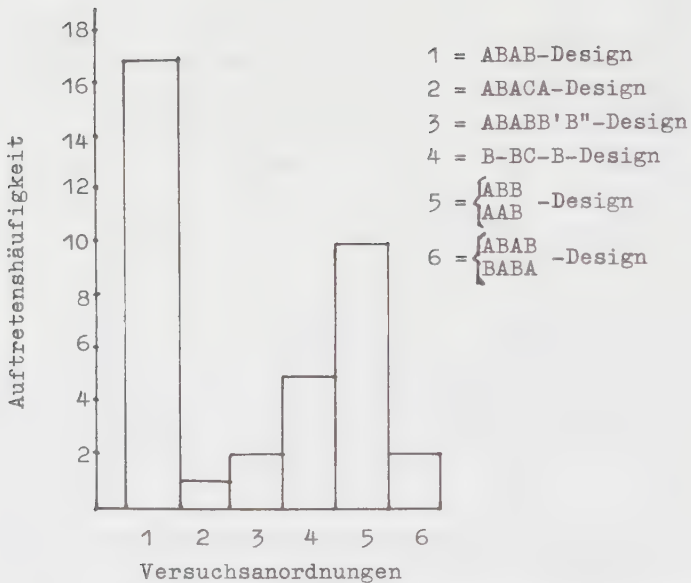


Abb. 5-2: Die Verteilung der Auftretenshäufigkeiten der verschiedenen Versuchsanordnungen der experimentellen Einzelfalluntersuchung bei den Untersuchungen des Schulleistungsverhaltens

Aus Abb. 5-1 ist zu ersehen, daß es nicht viele Kurse gibt, in denen das Schulleistungsverhalten mit der experimentellen Einzelfalluntersuchung untersucht wurde. Aber sie zeigt die Möglichkeit, daß die experimentelle Einzelfalluntersuchung auch für die Untersuchung des Schulleistungsverhaltens anwendbar ist. Abb. 5-2 und 5-3 zeigen, daß die Verteilung der Auftretenshäufigkeiten der verschiedenen Versuchsanordnungen der experimentellen Einzelfalluntersuchung für das Schulleistungsverhaltens (Abb. 5-2) und die Verteilung der Auftretenshäufigkeit der verschiedenen Beobachtungspunkte in einer Phase bei den Untersuchungen des Schulleistungsverhaltens (Abb. 5-3) starke Ähnlichkeit mit den Verteilungen haben, in denen alle abhängigen Variablen Zusammengefaßt sind (z.B. bei Abb. 5-2 und 4-4 sind das ABAB-

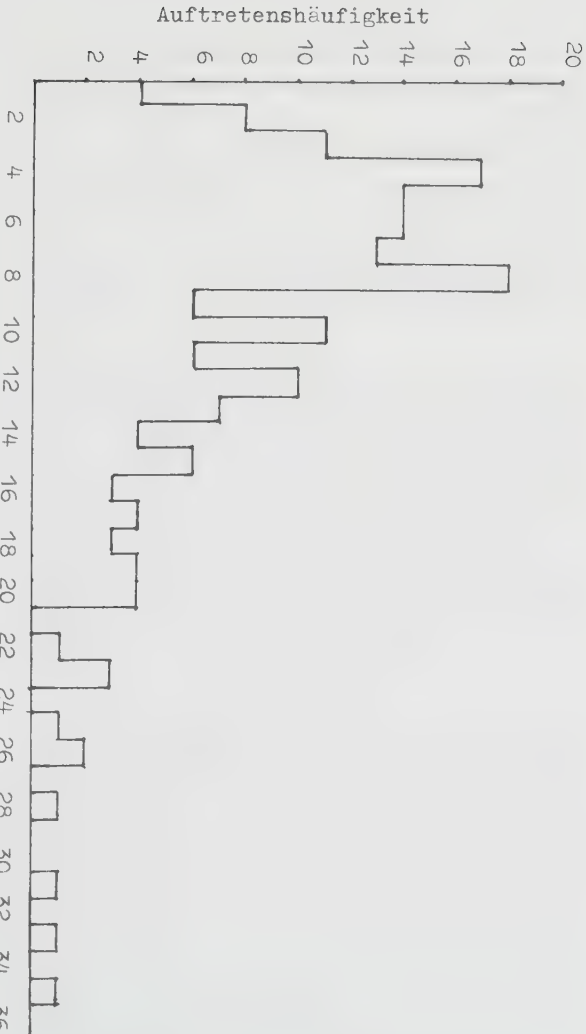


Abb. 5-3: Auftretenshäufigkeit jeder Phasenlänge bei den Untersuchungen der Schulleistungsverhaltens mit der experimentellen Einzel-falluntersuchung

und das Multiple-Baseline-Design am häufigsten verwendet worden, und bei Abb. 5-3 und 4-3 enthalten die meisten Phasen weniger als 30 Beobachtungen und die Phasenlängen von 4 bis 10 sind am stärksten vertreten), mit Ausnahme größerer Schwankungen, die durch kleinere Stichproben verursacht sind. Dies weist darauf hin, daß die Methodologie der experimentellen Einzelfalluntersuchung für die Erforschung des Schulleistungsverhalten nicht anders ist als die für andere abhängigen Variablen, mit Ausnahme der Datenerhebungsverfahren.

Wegen der unterschiedlichen Schwierigkeitsgrade der einzelnen Tests in der Untersuchung werden die Schwankungen der Kurve für Schulleistungsverhalten so groß wie die beim sozialen Verhalten. Wenn die Schwankungen der Kurve groß sind, muß die Phase verlängert werden.

5.3. Wie kann die experimentelle Einzelfalluntersuchung zur Theoriebildung herangezogen werden?

Um diese Fragestellung zu beantworten, müssen zunächst die Eigenschaften der Theorie, der Prozeß der Theoriebildung, und der Zusammenhang zwischen Theoriebildung und Versuchsanordnung erörtert werden. Erst dann kann man diskutieren, ob und in welcher Weise die experimentelle Einzelfalluntersuchung zur Theoriebildung herangezogen werden kann.

5.3.1. Die Eigenschaften der Theorie

Snow hat "Theorie" in ihrer einfachsten Form als eine symbolische Konstruktion definiert, die entworfen wird, um generalisierbare Fakten (oder Gesetzmäßigkeiten) in einen systematischen Zusammenhang zu bringen (Snow 1973, S.78).

Die formalen Eigenschaften einer Theorie werden im folgenden aufgeführt (Snow 1973, S.82-83):

Eine Theorie enthält

- (1) Axiome — dies sind deutlich dargelegte Postulate, aus denen empirisch überprüfbare Begriffe abgeleitet werden,
- (2) mathematische oder logische Formeln, die die Zusammenhänge zwischen Variablen festsetzen,
- (3) operationale Definitionen, die die in den von Postulaten abgeleiteten Theoremen enthaltenen Begriffe definieren,

(4) ein Modell, das die theoretische Konstruktion in verbalen, bildlicher, oder graphischer Form interpretiert.

Snow (1973) hat 12 Evaluationskriterien für die Angemessenheit einer Theorie aufgestellt. Sechs dieser Kriterien sind nützlich für die Verdeutlichung der Eigenschaften der Theorie. Dies sind folgende:

1. Eine Theorie soll ihre Postulate und die davon abgeleiteten Begriffe deutlich definieren.
2. Eine Theorie soll interne Konsistenz in ihrem logischen System haben.
3. Eine Theorie soll mit den bestehenden empirischen Daten übereinstimmen.
4. Eine Theorie soll spezifische Hypothesen und Vorhersagen hervorbringen können.
5. Eine Theorie soll falsifizierbar sein.
6. Eine Theorie soll quantifizierbar sein.

Snow (1973) klassifiziert Theorien hinsichtlich ihrer allgemeingültigen Wichtigkeit in sechs Grade (Theorie A-F):

(1) A-Theorie (Axiomatic Theory): Eine Theorie wird als axiomatische Theorie bezeichnet, wenn sie eine Gruppe von primitiven oder undefinierten Begriffen, mit deren Hilfe die übrigen Begriffe in der Theorie definiert werden können, und eine Gruppe von Postulaten (Axiomen), von denen die Theoreme abgeleitet werden können, enthält.

Diese Theorie bedarf umfassender mathematischer oder symbolisch-logischer Formulierungen, und ist nach Snow für erziehungswissenschaftliche Forscher unerreichbar, z.B. eine statistische Lerntheorie.

(2) B-Theorie (Broken Axiomatic Theory): Eine axiomatische Theorie kann dadurch verbessert werden, daß einige (nicht alle) ihrer Postulate in einer Untersuchung thematisiert, und die unangemessenen Postulate anhand der empirischen Daten, durch neu formulierte Postulate ersetzt werden. Man kann auch aus den alten Postulaten neue Formulierungen deduzieren. Angenommen, es gäbe beispielsweise vier axiomatische psychologische Theorien. Jede Theorie hätte Relevanz für die Erziehungswissenschaft. Die erziehungswissenschaftlichen Theoretiker könnten die nützlichen Teile jeder Theorie zur Re-Kombination für einen erzie-

hungswissenschaftlichen Zweck entnehmen.

- (3) C-Theorie (Conceptual theories and constructs): dies sind neue Konstrukte, die die Erklärung und Beschreibung der Beziehungen zwischen den beobachteten Variablen vereinfachen. Mit konsistenter Validation nimmt das Konstrukt eine Stelle in einem Netz von verbundenen Begriffen ein, die die Entdeckungen beschreiben, z.B. das Konstrukt "Intelligenz"; die Theorie der Leistungsmotivation (the theory of achievement motivation).
- (4) D-Theorie (Descriptive theories and taxonomies): Es handelt sich hierbei um Klassifikations- und Differenzierungsprozesse, die die beobachteten Phänomene systematisch beschreiben. Sie sind niedriger als C-Theorien klassifiziert worden, weil sie keine neuen hypothetischen Konstrukte für den Erklärungszweck einführen, z.B. Blooms Taxonomie der Lernziele.
- (5) E-Theorie (Elementisms): Die Phänomene werden durch logische oder empirische Verfahren in ihren Elementen oder Komponenten analysiert. Diese analysierten Elemente oder Komponenten können zur Entwicklung von D-Theorien synthetisiert werden, z.B. Guilfords Analyse der geistigen Leistungen; Arbeitsanalyse beim Berufstraining; und die Fertigungsanalyse beim Sozialfertigkeitstraining.
- (6) F-Theorie (Formative hypotheses): Sie beinhaltet Hypothesen für Problem-Lösungen oder Hypothesen für empirische Überprüfungen, z.B. die Pygmalion-Hypothese (Sie besagt, daß die günstigen Erwartungen des Lehrers auf das Kind den IQ-Zuwachs des Kindes fördern könnten. Diese Hypothese war Grundlage vieler Untersuchungen und auch der wahrscheinliche Mechanismus in den Erwartungen ist skizziert worden).

5.3.2. Prozesse der Theoriebildung

Obwohl Theorien jeden Grades im allgemeinen bestimmte Zusammenhänge zwischen Begriffen und empirischen Daten postulieren, sind die Prozesse der Theoriebildung nach Snow für jeden Grad unterschiedlich. Theorien niedrigeren Grades (D-, E-, und F-Theorie) werden durch die relativ einfache Zusammenfassung von empirischen Zusammenhängen

ohne substantielle Schlüsse oder deduktive Logik charakterisiert. Der Schwerpunkt liegt in dem Erwerb der Fakten. Nachdem die Fakten gesammelt worden sind, werden generalisierte Differenzierungen, Kategorien oder Prinzipien induziert. Eine Theorie von Grad C oder vielleicht B bedeutet eine kontinuierliche Interaktion zwischen den vorläufigen theoretischen Begriffen und der Datensammlung. Der Forscher formuliert ein neues Konstrukt, um die Erfindung darzustellen. Zur Überprüfung der Validität dieses neuen Konstrukts werden oft weitere empirische Untersuchungen durchgeführt. Manchmal muß das neue Konstrukt aufgrund der empirischen Ergebnisse revidiert werden. Die A-Theorie ist eine deduktive Theorie. Die Lücke zwischen der Theorie und der empirischen Untersuchung ist bei der A-Theorie größer als bei den Theorien niedrigeren Grades. Die Theorie ist aus einer Mehrzahl von Daten formal konstruiert worden. Ihre mathematische und logisch-deduktive Struktur ist sorgfältig. Um die abgeleiteten Hypothesen zu überprüfen, ist ein Programm von Untersuchungen erforderlich, und die Ergebnisse führen gewöhnlich zur Verbesserung der Theorie.

Für die Prozesse der Theoriebildung nennt Snow (1973) ein MMM-Kontinuum. Es ist: 1. Anerkennung der Metaphern, die die wichtigsten Aspekte der Phänomene einfangen, 2. Kodifizierung der Modelle: Die Grundzüge der Metaphern müssen in eine formale Repräsentation (Modell) übersetzt oder kodifiziert werden, 3. Organisation von Metatheorien: Theoretische Modelle werden aufgrund kontinuierlicher Interaktion mit empirischen Beobachtungen und Datenanalysen organisiert, modifiziert und eventuell wieder organisiert. Die organisierten Modelle heißen Metatheorien.

5.3.3. Die Bedeutung der experimentellen Einzelfalluntersuchung bei der Theoriebildung

Die obigen kurzen Skizzen über die Eigenschaften der Theorie und die Prozesse der Theoriebildung in Anlehnung an Snows "Reviewartikel" weisen darauf hin, daß die Theoriebildung nichts mit der Art der Versuchsanordnung zu tun hat. D.h. eine beliebige Versuchsanordnung kann zur Theoriebildung herangezogen werden, sofern sie die von der Theorie ab-

geleiteten Hypothesen überprüfen und valide Aussagen liefern kann. Es ist gleichgültig, ob es sich dabei um Faktorial-Kontrollgruppen- oder Einzelfall-Experimentaldesign handelt. Nach Tabelle 2-2 werden die gegenwärtigen erziehungswissenschaftlichen Forschungsmethoden von Faktorial- und Kontrollgruppensdesign dominiert. Diese beiden Versuchsanordnungen bedürfen einer großen Anzahl von Versuchspersonen, um durch Randomisierungs- bzw. Parallelisierungsverfahren die Gleichheit der Ausgangslage der zu vergleichenden Gruppen zu gewährleisten. Je größer die Anzahl der Gruppen ist, desto schwieriger ist es, die Zusammenarbeit der Schulverwaltung und der Lehrer zu erreichen. Die experimentellen Einzelfalluntersuchungen brauchen relativ wenige Vpn und sind deshalb leichter anwendbar als die Gruppensdesigns.

Wenn man bei der Überprüfung der Hypothese nur einen Faktor (Treatment) manipuliert, ist die experimentelle Einzelfalluntersuchung sehr angemessen. Aber wenn mehrere Faktoren manipuliert werden, ist die experimentelle Einzelfalluntersuchung der Überprüfung der Wechselwirkungen zwischen den Faktoren nicht gewachsen. In diesem Fall ist eine koordinierende Untersuchung mit dem faktoriellen Design zu empfehlen. Z.B. liegt in den Prozessen der direkten und systematischen Wiederholungen wahrscheinlich eine Wechselwirkung des Haupttreatments mit einer Persönlichkeitsvariablen vor, wird es effizienter sein, ein faktorielles Design mit einzubeziehen.

Bisher ist die experimentelle Einzelfalluntersuchung vorwiegend im Bereich der Experimentalanalyse der beobachtbaren Interaktion des Organismus mit den Ereignissen in seiner Umwelt angewandt worden. Obwohl Bijou (1970) meint, daß diese Untersuchungen nicht zur Überprüfung einer Theorie oder einer Hypothese, sondern zur Evaluierung einer praktischen Anwendung der Begriffe und Prinzipien geplant werden, überprüft die experimentelle Einzelfalluntersuchung doch, im Sinn von Snows Klassifikation der Theorien, die von Theorien abgeleiteten Hypothesen. Die überprüften Theorien können von der F-Theorie (z.B. die einfachste Hypothese für die Lösung des Problems eines behinderten Kindes) bis zur C-Theorie (z.B. neue Konstrukte in der Verhaltenswissenschaft) reichen. Es

hängt davon ab, von welchem Hintergrund die Hypothese abgeleitet wird. Wenn ein Konzept mehrere empirische Untersuchungen anregen kann, ist es doch eine wertvolle Theorie, obgleich es nur eine Theorie niedrigeren Grades ist, z.B. das Konzept der "Personalisierten Instruktion". Dieses Konzept ist zuerst von Skinner in Form der "Programmierten Instruktion" (oder teaching mashine) realisiert, von Bloom (z.B. Bloom 1970) auf die schulische Ebene und von Keller (1968) auf die hochschulische Ebene übertragen worden, und hat schon empirische Untersuchungen angeregt (z.B. Semb et al. 1973; Miller & Weaver 1976; Semb 1974; Miller et al. 1974).

Nach König ist "Erziehungswissenschaft nur dann sinnvoll, wenn sie letztlich ausmündet in Handlungsanweisungen für die Erziehungspraxis" (König 1978, S.28). Er fordert, daß die empirische Erziehungswissenschaft generelle Aussagen über Ursachen und Wirkungen verschiedener Handlungen bzw. verschiedener Situationen in der Erziehungspraxis aufstellen soll (vgl. König 1978, S.103), und er weist darauf hin, daß die empirische Erziehungswissenschaft eine doppelte Zeilsetzung haben soll, d.h. die wissenschaftliche Überprüfbarkeit und die Praktikabilität (vgl. König 1978, S.135). Die experimentelle Einzelfalluntersuchung kann die von König (1978) aufgestellten Forderungen für die erziehungswissenschaftliche Forschung erfüllen. Sie ist einerseits wissenschaftlich, seine interne und externe Validität sind durch verschiedene Verfahren erreichbar (siehe Abschnitt 3 der vorliegenden Arbeit), und andererseits ist sie in konkreten Schulsituationen praktikabel (siehe Abb.4-6). Abb. 5-4 postuliert die Stelle und Funktion der experimentellen Einzelfalluntersuchung im pädagogischen Theorie-Praxis-Verhältnis. Untersuchung II der vorliegenden Arbeit zeigt, daß die experimentelle Einzelfalluntersuchung für die Erforschung des sozialen und Schulleistungsverhaltens in Schulsituationen mit fruchtbaren Ergebnissen eingesetzt werden kann. Weitere Fortschritte in der grundlegenden und technologischen Kenntnis sollten natürlich zur effektiven Anwendung führen. Zum Beispiel:
Bei der Schul- oder Unterrichtsreform kann beispielsweise

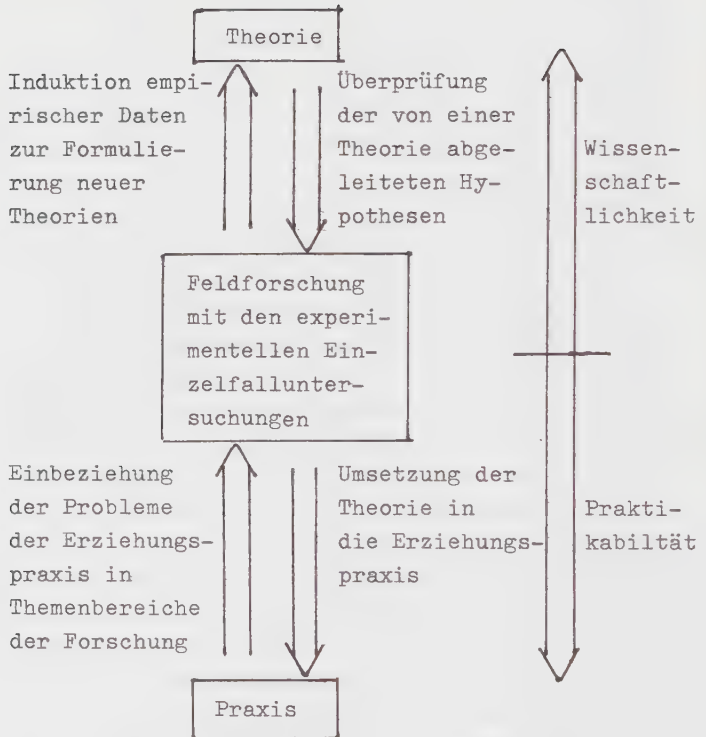


Abb. 5-4: Modell über die Stelle und Funktion der experimentellen Einzelfalluntersuchungen im pädagogischen Theorie-Praxis-Verhältnis

das Multiple-Baseline-Design zur Evaluation des Reformerfolgs herangezogen werden, wenn die zu überprüfenden Hypothesen, die aus erziehungswissenschaftlicher Theorie abgeleitet sind, und die Indikatoren des Reformerfolgs (abhängige Variablen) operational definiert werden (vgl. Krumm 1973). Die Versuchsanordnung sieht wie folgendes aus:

		Schulische Einheiten			
		1	2	3	
Phasen	1	A	A	A	wobei
	2	B	A	A	A = Ohne Einsatz
	3	B	B	A	der Reform
	4	B	B	B	B = Einsatz der Reform

Die schulische Einheit kann eine Klasse, eine Schule oder ein Schuldistrikt sein. In der ersten Phase werden von allen in die Untersuchung einbezogenen schulischen Einheiten Baseline-Daten erhoben. In der zweiten Phase wird der Einsatz der Reform auf Schuleinheit 1 beschränkt. In der dritten Phase wird die Reform auf die weitere Schuleinheit eingesetzt, usf. Bei dieser Untersuchung handelt es sich um die Überprüfung der Anwendbarkeit einer Theorie in die Erziehungspraxis; also geht die Feldforschung mit den experimentellen Einzelfalluntersuchungen in die Richtung von oben nach unten bei Abb. 5-4. Wenn bei der Schulreform die unabhängigen Variablen eindeutig operationalisiert werden, wie es von Krumm gefordert wird (Krumm ist der Ansicht, daß die Reformschritte so klein wie möglich sein soll, damit die tatsächlich funktionierenden Faktoren identifiziert werden können, siehe Krumm 1973, S.263), dann können die generellen Aussagen über Ursachen und Wirkungen der erziehungswissenschaftlichen Handlung aufgestellt werden. Dies kann zur Entwicklung der erziehungswissenschaftlichen Erkenntnis beitragen. In diesem Fall geht die Feldforschung mit den experimentellen Einzelfalluntersuchungen in die Richtung von unten nach oben bei Abb. 5-4.

6. Zusammenfassung

Die vorliegende Arbeit überprüft die Anwendbarkeit der experimentellen Einzelfalluntersuchung für die erziehungswissenschaftliche Feldforschung. Um die gegenwärtige erziehungswissenschaftliche Forschungslage in der BRD zu verstehen, wurden 77 experimentelle Artikel aus vier erziehungswissenschaftlichen Zeitschriften analysiert. Die Ergebnisse zeigen, daß a) die gegenwärtige erziehungswissenschaftlich experimentelle Forschung in der BRD viel mehr an der Feldforschung als an der Laborforschung orientiert ist, und b) die experimentelle Einzelfalluntersuchung noch nicht von den Erziehungswissenschaftlern in der BRD über die Verhaltensmodifikationsforschung hinaus als ein Instrument für die Feldforschung akzeptiert worden ist. Um die interne und externe Validität der experimentellen Einzelfalluntersuchung zu untersuchen, wurden insgesamt 367 Artikel als Stichproben aus vier Zeitschriften zur Verhaltensmodifikation gezogen. Wichtige Ergebnisse sind folgende: 1) Nur ca. ein Drittel der Phasenpaare erreicht das Kriterium des Nicht-Überschneidens. Dies besagt, daß statistische Verfahren zur Überprüfung der Phasendifferenz erforderlich sind. 2) Ein durchschnittlicher Prozentsatz der Beobachterübereinstimmung von 93.6% mit $\sigma = 0.85\%$ liegt ohne große Schwankungen bei verschiedenen Zeitschriften und Jahrgängen vor. Dieser Prozentsatz kann als Kriterium für Beobachtertraining dienen. 3) Zur Überprüfung des Behandlungseffekts sind bei N=1-Design Zeitreihenanalyse und visuelle Beurteilung anwendbar. Bei N=1-Gruppe-Design stehen einige Prüfverfahren zur Verfügung (z.B. t-Test für korrelierte Stichproben, Varianzanalyse für wiederholte Messungen). Fazit: die experimentelle Einzelfalluntersuchung ist in der erziehungswissenschaftlichen Feldforschung anwendbar. Außerdem kann der Praktiker (z.B. ein Lehrer), wenn er mit der experimentellen Einzelfalluntersuchung vertraut ist, die Schwierigkeit eines einzelnen Kinds mit N=1-Design untersuchen. Dadurch hilft er nicht nur dem Kind bei seinem Verhaltensproblem, er trägt auch zur Erforschung des Mechanismus der Verhaltensänderung bei. Er kann auch in seiner Klasse eigene Unterrichtsforschung mit N=1-Gruppe-Design betreiben.

Literatur

- Arnold, S.C., Forehand, R., and Trurgis, E.T.: Effects of a response-cost procedure on the academic performance of retarded students. Journal of Behavior Therapy and Experimental Psychiatry, 1976, 7, 191-192.
- Barlow, D.H. and Hersen, M.: Single-case experimental designs: Uses in applied clinical research. Archives of General Psychiatry, 1973, 29, 319-325.
- Barlow, D.H., Agras, W.S., Leitenber, H., Callahan, E.J. and Moore, R.C.: Case histories and shorter communications. Behavior, Research and Therapy, 1972, 10, 411-415.
- Bijou, S.W. What psychology has to offer education — now. Journal of Applied Behavior Analysis, 1970, 3, 65-71.
- Bijou, S.W., Peterson, R.F., Harris, F.R., Allen, K.E. and Johnston, M.S.: Methodology for experimental studies of young children in natural settings. The Psychological Record, 1969, 19, 177-210.
- Bijou, S.W., Peterson, R.F. and Ault, M.H.: A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. Journal of Applied Behavior Analysis, 1968, 1, 175-191.
- Bloom, D.S.: Alle Schüler schaffen es. Betrifft: Erziehung 1970, 11, 15-27.
- Böttger, P., Bastine, R., Kirchner, F.Th. und Kissel, E.: Zum Informationsinhalt verhaltenstherapeutischer Veröffentlichungen — eine empirische Analyse. In: F. Petermann (Hrsg.): Psychotherapieforschung. Weinheim und Basel, 1977, 103-140.
- Boudewyns, P.A.: A comparison of the effects of stress vs. relaxation instruction on the finger temperature response. Behavior Therapy, 1976, 7, 54-67.
- Box, G.E. & Jenkins, G.M.: Time-series analysis: forecasting and control. San Francisco: Holden Day, 1970.
- Bracht, G.H.: Experimentelle Faktoren in Beziehung zur Wechselwirkung zwischen Schülermerkmalen und Unterrichtsmethoden. In: R. Schwarzer & K. Steinhagen (Hrsg.): Adaptiver Unterricht. München: Kosel, 1975, S.94-108.
- Bracht, G.H. und Glass, G.V.: The external validity of

- experiments. American Educational Research Journal 1968, S.437-474. Die deutsche Übersetzung: Die externe Validität von Experimenten. In: R. Schwarzer und K. Seinhagen (Hrsg.): Adaptiver Unterricht. München:Kösel, 1975, S.64-93.
- Brezinka, W.: Von der Pädagogik zur Erziehungswissenschaft. Weinheim 1971.
- Brezinka, W.: Die Pädagogik und die erzieherische Wirklichkeit. In: H. Möhrs: Erziehungswissenschaft und Erziehungswirklichkeit. Frankfurt: Akademische Verlagsgesellschaft, 1964.
- Brady, J.P., Luborsky, L. and Kron, R.E.: Blood pressure reduction in patients with essential hypertension through metronome-conditioned relaxation: A preliminary report. Behavior Therapy, 1974, 5, 203-209.
- Calhoun, K.S. and Matherne, P.: The effects of varying schedules of time-out on aggressive behavior of a retarded girl. Journal of Behavior Therapy and Experimental Psychiatry, 1975, 6, 139-143.
- Campbell, D.T. and Stanley, J.C.: Experimental and quasi-experimental designs for research on teaching. In: N.L. Gage (Hrsg.): Handbook of research on teaching. Chicago, 1963. Die deutsche Übersetzung : Experimentelle und quasi-experimentelle Anordnungen in der Unterrichtsforschung. In: K. Ingenkamp (Hrsg.): Strategien der Unterrichtsforschung. Weinheim und Basel: Beltz 1973, S.99-193. oder in: K. Ingenkamp und E. Parey (Hrsg.): Handbuch der Unterrichtsforschung. Teil I. Weinheim: Beltz, 1970.
- Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960, 20, 37-46.
- Creer, T.L. & Davis, M.H.: Using a staggered-wakening procedure with enuretic children in an institutional setting. Journal of Behavior Therapy and Experimental Psychiatry, 1975, 6, 23-25.
- Cronbach, L.J.: Wie kann Unterricht an individuelle Unterschiede angepaßt werden? In: R. Schwarzer & K. Steinhagen (Hrsg.): Adaptiver Unterricht. München: Kösel, 1975a, S.42-58.
- Cronbach, L.J.: Beyond the two disciplines of scientific Psychology. American Psychologist, 1975b, 30, 116-127.

- Cronbach, L.J. & Snow, R.E.: Aptitudes and Instructional methods: a handbook for research on interaction. New York: Irvington 1977.
- Curran, J.P., Lentz, R.J., and Paul, G.L.: Effectiveness of Sampling-Exposure Procedure on facilities utilization by psychiatric head-core chronic patients. Journal of Behavior Therapy and Experimental Psychiatry, 1973, 4, 201-207.
- DFG-Kommission Erziehungswissenschaft: Erziehungswissenschaftliche Schwerpunktprogramme der deutschen Forschungsgemeinschaft. Zeitschrift für Pädagogik, 1974, 20, 967-972.
- Doty, D.W., McInnis, T., and Paul, G.L.: Remediation of negative side effects of an on-going response-cost system with chronic mental patients. Journal of Applied Behavior Analysis, 1974, 7, 197-198.
- Drabman, R.S. and Lahey, B.B.: Feedback in classroom behavior modification: effects on the target and her classmates. Journal of Applied Behavior Analysis, 1974, 7, 591-598.
- Edward, A.L.: Versuchsplanung in der psychologischen Forschung. Weinheim, Berlin, Basel: Beltz, 1971.
- Epstein, L.H., Parker, L., McCoy, J.F., and McGee, G.: Descriptive analysis of eating regulation in obese and nonobese children. Journal of Applied Behavior Analysis, 1976, 9, 407-415.
- Epstein, L.H., Hersen, M., and Hemphill, D.P.: Music feedback in the treatment of tension headache: An experimental case study. Journal of Behavior Therapy and Experimental Psychiatry, 1974, 5, 59-63.
- Eriksson, J.H., Götestam, K.G., Melin, L. and Öst, L.G.: A token economy treatment of drug addiction. Behavior, Research and Therapy, 1975, 13, 113-125.
- Feingold, B.D. and Mahoney, M.J.: Reinforcement effects on intrinsic interest: undermining the overjustification hypothesis. Behavior Therapy, 1975, 6, 367-377.
- Fischer, A.: Über die Bedeutung des Experiments in der pädagogischen Forschung und die Idee einer exakten Pädagogik. In: H. Röhrs: Erziehungswissenschaft und Erziehungswirksamkeit. Frankfurt: Akademische Verlagsgesellschaft, 1964.
- Flehsig, K.H.: Die Funktion des Experiments in der Unter-

- richtsforschung. Die Deutsche Schule, 1967, 59, 397-413.
- Gadenne, V.: Die Gültigkeit psychologischer Untersuchung. Stuttgart: Kohlhammer, 1976.
- Garrett, H.E.: Statistics in Psychology and Education. 5. Ed. London: Longmans, 1965.
- Gentile, J.R., Roden, A.H., and Klein, R.D.: An Analysis of Variance model for the intrasubject replication design. Journal of Applied Behavior Analysis, 1972, 5, 193-198.
- Glass, G.V., Willson, V.L., and Gottman, J.M.: Designs and analysis of time-series experiments. Boulder, Colorado: Colorado Associated University Press, 1975.
- Glover, J. and Gary, A.L.: Procedures to increase some aspects of creativity. Journal of Applied Behavior Analysis, 1976, 9, 79-84.
- Glynn, E.L. and Thomas, J.D.: Effect of cueing on self-control of classroom behavior. Journal of Applied Behavior Analysis, 1974, 7, 299-306.
- Haerberlin, U.: Empirische Analyse und pädagogische Handlungsforschung. Zeitschrift für Pädagogik, 1975, 21, 653-676.
- Hagemann, W.: Paradigmatische Probleme der experimentellen Unterrichtsforschung. Zeitschrift für erziehungswissenschaftliche Forschung, 1976, 10, 127-144.
- Hartmann, D.P.: Considerations in the choice of interobserver reliability estimates. Journal of Applied Behavior Analysis, 1977, 10, 103-117.
- Hauserman, N., Walen, S.R., and Behling, M.: Reinforced racial integration in the first grade: a study in generalization. Journal of Applied Behavior Analysis, 1973, 6, 193-200.
- Hay, L.R., Nelson, R.O., and Hay, W.M.: The use of teachers as behavioral observers. Journal of Applied Behavior Analysis, 1977, 10, 345-348.
- Hays, W.L.: Statistics. New York: Holt, Rinehart and Winston Inc. 1969.
- Hehlmann, W.: Wörterbuch der Psychologie. Stuttgart: Alfred Kröner, 1974.
- Heidenreich, W.D. und Heymann, H.W.: Lehr-Lern-Forschung. Zeitschrift für Pädagogik, 1976, 22, 225-251.
- Heinze, U., Müller, B.S., und Zinnecker, J.: Handlungsforschung im pädagogischen Feld. München: Juventa, 1975.

- Hersen, M. and Barlow, D.: Single case experimental designs: strategies for studying behavior change. New York: Pergamon Press, 1976.
- Hollander, M.A. and Plutchik, R.: A reinforcement program for psychiatric attendants. Journal of Behavior Therapy and Experimental Psychiatry, 1972, 3, 297-300.
- Holzkamp, A.E.: Kritische Psychologie. Frankfurt am Main, 1972.
- Huber, H.P.: Kontrollierte Fallstudie. In: L.J. Pongratz (Hrsg.): Handbuch der Psychologie. Bd. 3, 2. Klinische Psychologie. Göttingen: Hogrefe, 1978.
- Jones, F.H. and Eimers, R.C.: Role playing to train elementary teachers to use a classroom management skill package. Journal of Applied Behavior Analysis, 1975, 8, 421-433.
- Jones, R.R., Weinrott, M.R. and Vaught, R.S.: Effect of serial dependency on the agreement between visual and statistical inference. Journal of Applied Behavior Analysis, 1978, 11, 277-283.
- Jones, R.R., Vaught, R.S. and Weinrott, M.R.: Time-series analysis in operant research. Journal of Applied Behavior Analysis, 1977, 10, 151-166.
- Johnson, S.M. and Bolstad, O.D.: Reactivity to home observation: a comparison of audio-recorded behavior with observers present or absent. Journal of Applied Behavior Analysis, 1975, 8, 181-185.
- Kaufman, K.M. and O'Leary, K.D.: Reward, cost, and self-evaluation procedures for disruptive adolescents in a psychiatric hospital school. Journal of Applied Behavior Analysis, 1972, 5, 293-309.
- Kazdin, A.E. and Klock, J.: The effect of nonverbal teacher approval on student attentive behavior. Journal of Applied Behavior Analysis, 1973, 6, 643-654.
- Kazdin, A.E.: The Impact of Applied behavior analysis on diverse area of research. Journal of Applied Behavior Analysis, 1975, 8, 213-229.
- Kazdin, A.E.: Statistical analyses for single-case experimental designs. In: M. Hersen and D. Barlow: Single case experimental designs: Strategies for studying behavioral change. New York: Pergamon Press, 1976, S. 265-316.

- Kazdin, A.E.: Artifact, bias and complexity of assessment. Journal of Applied Behavior Analysis, 1977, 10, 141-150.
- Keller, F.S.: "Good-bye teacher...". Journal of Applied Behavior Analysis, 1968, 1, 79-89.
- Kendall, P.C., Nay, W.R. and Jeffers, J.: Time-out duration and contrast effects: a systematic evaluation of a successive treatments design. Behavior Therapy, 1975, 6, 609-615.
- Kent, R.N., Kanowitz, J., O'Leary, D. and Cheiken, M.: Observer reliability as a function of circumstances of assessment. Journal of Applied Behavior Analysis, 1977, 10, 317-324.
- Kerlinger, F.N.: Foundation of behavioral research: educational and psychological inquiry. London: Holt, Rinehart & Winston, 1972.
- Kirchner, F.T., Kissel, E., Petermann, F. und Böttger, P.: Interne und externe Validität empirischer Untersuchungen in der Psychotherapieforschung. In: F. Petermann (Hrsg.): Psychotherapieforschung. Weinheim und Basel: Beltz, 1977, S.51-102.
- Köck, P und Ott, H.: Wörterbuch für Erziehung und Unterricht. Donauwörth: Ludwig Auer, 1976.
- König, E.: Theorie der Erziehungswissenschaft. Band 3. München: Wilhelm Fink Verlag. 1978.
- Kratochwill, T.R. and Wetzell, R.J.: Observer agreement, credibility and judgement: some considerations in presenting observer agreement data. Journal of Applied Behavior Analysis, 1977, 10, 133-139.
- Krumm, V.: Experimentelle Einzelfallforschung — eine notwendige Ergänzung der klassischen Unterrichtsforschung? Vortrag in der 25. Tagung der AEPF (Arbeitsgruppe für Empirische Pädagogische Forschung) von November 1977 in Düsseldorf, nicht veröffentlicht.
- Krumm, V.: Die Rolle der Erziehungswissenschaft bei der Reform von Schule und Unterricht. In: Jahrbuch der Universität Düsseldorf, 1972/73. Düsseldorf: Triltsch, 1973 S.251-265.
- Krumm, V.(Hrsg.): Zur Handlungsrelevanz der Verhaltenstheorien. München. Wien. Baltimore: Urban & Schwarzenberg, 1979.
- Kubany, E.S., Weiss, L.E. and Sloggett, B.B.: The good behavior clock: a reinforcement/time-out procedure for re-

- ducing disruptive classroom behavior. Journal of Behavior Therapy and Experimental Psychiatry, 1971, 2, 173-179.
- Lanyon, R.I., Barrington, C.C. and Newman, A.C.: Modification of stuttering through EMG biofeedback: a preliminary study. Behavior Therapy, 1976, 7, 96-103.
- Leitenberg, H., Wincze, J.P., Butz, R.A. and Callahan, E.J.: Comparison of the effects of instructions and reinforcement in the treatment of a neurotic avoidance response: a single case experiment. Journal of Behavior Therapy and Experimental Psychiatry, 1970, 1, 53-58.
- Lewis, D.G.: Experimentelle Planung in der Erziehungswissenschaft. Weinheim & Basel: Beltz, 1974.
- Lochner, R.: Deutsche Erziehungswissenschaft. Meinsenhein: Hain, 1963.
- Mackie, R.R. and Christensen, P.R.: Translation and application of psychological research. Technical Report, 716-1, January, 1967. Human Factors Research Incorporated, Santa Barbara California.
- McInnis, T., Himelstein, H.C., Doty, D.W. and Paul, G.L.: Modification of sampling-exposure procedures for increasing facilities utilization by chronic psychiatric patients. Journal of Behavior Therapy and Experimental Psychiatry, 1974, 5, 119-127.
- Miller, L.K. and Weaver, F.H.: A behavioral technology for producing concept formation in university students. Journal of Applied Behavior Analysis, 1976, 9, 289-300.
- Miller, L.K., Weaver, F.H. and Semb, G.: A procedure for maintaining student progress in a personalized university course. Journal of Applied Behavior Analysis, 1974, 7, 87-91.
- Myers, J.L.: Foundamentals of experimental design. Boston: Allyn & Bacon, 1972.
- O'Brien, F., Azrin, N.H. and Herson, K.: Increased communications of chronic mental patients by reinforcement and by response priming. Journal of Applied Behavior Analysis. 1969, 2, 23-29.
- C'Leary, K.D., Becker, W.C., Evans, M.B. and Saudargas, R.A.: A token reinforcement program in a public school: a replication and systematic analysis. Journal of Applied

- Behavior Analysis, 1969, 2, 3-13.
- Petermann, F.(Hrsg.): Psychotherapieforschung. Weinheim: Beltz, 1977a.
- Petermann, F.(Hrsg.): Methodische Grundlagen klinischer Psychologie. Weinheim: Beltz, 1977b.
- Petermann, F.: Veränderungsmessung. Stuttgart:Kohlmannner, 1978.
- Petermann, F. und Hehl F.(Hrsg.): Einzelfallanalyse. München-Wien-Baltimore: Urban & Schwarzenberg, 1979.
- Powell, J., Martindale, A. and Kulp, S.: An evaluation of time-sample measures of behavior. Journal of Applied Behavior Analysis, 1975, 8, 463-469.
- Reif, F.: Toward an applied science of Education: some key questions and directions. Instructional Science, 1978, 7, 1-14.
- Renne, C.M. and Creer, T.L.: Training children with asthma to use inhalation therapy equipment. Journal of Applied Behavior Analysis, 1976, 9, 1-11.
- Repp, A.C., Roberts, D.M., Slack, D.J., Repp, C.F. and Berkler, M.S.: A comparison of frequency, interval and time-sampling methods of data collection. Journal of Applied Behavior Analysis, 1976, 9, 501-508.
- Richard, H.C. and Saunders, Jr. T.R.: Control of "clean-up" behavior in a summer camp. Behavior Therapy, 1971, 2, 340-344.
- Romanczyk, R.G., Kent, R.N., Diamant, C., and O'Leary, K.D.: Measuring the reliability of observational data: a reactive process. Journal of Applied Behavior Analysis, 1973, 6, 175-184.
- Roth, H. und Friedrich, D.(Hrsg.): Bildungsforschung: Problem-Perspektiven-Prioritäten. Teil I. Stuttgart, 1975.
- Rubin, B.K. and Stolz. S.B.: Generalization of self-referent speech established in a retarded adolescent by operant procedures. Behavior Therapy, 1974, 5, 93-106.
- Sachs, D.A., Martin, J.E. and Fitch, J.L.: The effects of visual feedback on a digital exercise in a functionally deaf cerebral palsied child. Journal of Behavior Therapy and Experimental Psychiatry, 1972, 3, 217-222.
- Scheffé, H.: The analysis of variance. New York: John Wiley

& Sons, Inc. 1959.

- Schnelle, J.F., Kirchner, R.E., McNeese, M.P. and Lawler, J.M.: Social evaluation research: the evaluation of two police patrolling strategies. Journal of Applied Behavior Analysis, 1975, 8, 353-365.
- Schoenfeld, W.N., Cumming, W.W. and Hearst, E.: On the classification of reinforcement schedules. Proc. Nat. Acad. Sci., 1956, 42, S.567. Zitiert von Sidman, M.: Tactics of scientific research: Evaluating experimental data in psychology. New York, Basic Books, 1960, S.260.
- Schulman, L.S.: Pädagogische Forschung — Versuch einer Neufassung. Die Deutsche Schule, 1974, 66, 798-809 (Teil I); 1975, 67, 6-17 (Teil II).
- Semb, G., Hopkins, B.L. and Hursh, D.E.: The effects of study questions and grades on students test performance in a college course. Journal of Applied Behavior Analysis, 1973, 6, 631-642.
- Semb, G.: The effects of mastery criteria and assignment length on college-student test performance. Journal of Applied Behavior Analysis, 1974, 7, 61-69.
- Shine, L.C. and Bower, S.M.: A one-way analysis of variance for single-subject designs. Educational and Psychological Measurement, 1971, 31, 105-113.
- Sidman, M.: Tactics of scientific research: Evaluating experimental data in psychology. New York: Basic Books, 1960.
- Siegel, S.: Nichtparametrische statistische Methoden. Frankfurt am Main: Fachbuchhandlung für Psychologie, 1976.
- Snow, R.E.: Theory construction for research on teaching. In: R.M.W. Travers: Second Handbook of research on teaching. Chicago: Rand McNally College Publishing Company, 1973.
- Snow, R.E.: Representative and Quasi-Representative designs for research on teaching. Review of Educational Research, 1974, S.265-291.
- Thomas, D.R., Becker, W.C. and Armstrong, M.: Production and elimination on disruptive classroom behavior by systematically varying teacher's behavior. Journal of Applied Behavior Analysis, 1968, 1 35-45.
- Travers, R.M.W.: Einführung in die erziehungswissenschaftliche Forschung. München, Oldenbourg, 1972.

Tunner, W. und Florin, I.: Kontrollen im Verhaltenstherapeutischen Experiment, Mitteilungen der Gesellschaft zur Förderung der Verhaltenstherapie, Heft 1, 1969. Zitiert von Hartig, M.: Problem und Methoden der Psychotherapieforschung. Urban & Schwarzenberg, München-Berlin-Wien, 1975, S.63.

Werres, W.: Empirische Erziehungswissenschaft in ihrer Bedeutung für die Theorie des Unterrichts. Essen: Neue Deutsche Schule, 1975.

Winkler, R.C.: Management of chronic psychiatric patients by a token reinforcement system. Journal of Applied Behavior Analysis, 1970, 3, 47-55.

Anhang 1: Seitenangaben der ausgewählten Artikel für
Untersuchung I

Zeit- schriften	<u>PEU</u>	<u>PEU</u>	<u>ZEPP</u>	<u>ZeF</u>	<u>DS</u>
Seiten-	(1972)	110-115	(1972)	(1974)	(1972)
nummer	46-54	121-125	1-16	82-103	362-376
der	80-88	131-136	105-117	(1975)	398-408
Artikel	88-98	144-152	151-164	164-180	477-486
	227-233	173-176	221-234		(1973)
	257-268	215-223	249-260		224-234
	268-277	242-245	(1973)		(1974)
	300-309	246-249	1-15		838-850
	379-384	283-288	16-28		(1976)
	(1973)	328-336	62-68		766-774
	14-30	351-354	91-105		
	77-88	355-363	167-181		
	156-162		253-262		
	223-234		279-292		
	235-243		293-302		
	279-288		(1974)		
	297-307		75-89		
	327-336		109-123		
	(1974)		192-206		
	15-21		225-240		
	22-34		262-280		
	35-43		(1975)		
	164-175		73-87		
	249-253		203-213		
	293-302		(1976)		
	303-309		37-43		
	(1975)		44-50		
	129-136		58-61		
	167-171		117-134		
	284-294		135-141		
	(1976)		220-223		
	21-28		235-244		
	52-59		274-287		
			298-308		

Anhang 2: Seitenangaben der ausgewählten Artikel, in denen die unabhängigen Variablen untersucht worden sind (für Untersuchung I).

1. Belastungssituation

FEU: 1973 (14-30); (235-243); 1974 (293-302); 1976 (246-249).

ZEPP: 1973 (253-262); (293-302).

2. Einstellung

FEU: 1976 (121-125).

ZEPP: 1976 (298-308).

3. Lernhilfe

FEU: 1976 (173-176).

ZEPP: 1974 (225-240).

4. Modell-Lernen

SP (Die Zeitschrift von "Psychologie in Erziehung und Unterricht" bevor Jahrgang 1972 heißt "Schule und Psychologie): 1972 (300-309); FEU: 1974 (164-175).

ZEPP: 1972 (221-234); 1974 (75-89); (262-280); 1975 (73-87).

5. Persönlichkeitsvariablen

FEU: 1973 (14-30); 1976 (21-28); (173-176); (215-223); (242-245); (355-363).

ZEPP: 1972 (1-16); (151-164); 1973 (1-15); (62-68); (167-181); (293-302); 1974 (192-206); 1975 (73-87); 1976 (117-134); (135-141); (235-244).

6. Schulform

SP: 1972 (268-277); FEU: 1973 (297-307); 1976 (144-152).

DS: 1974 (838-850).

7. Unterrichtsform

FEU: 1973 (156-162); (327-336); 1975 (284-294); 1976 (21-28); (131-136); (242-245); (351-354); (355-363).

ZEPP: 1976 (37-43); (135-141).

DS: 1972 (362-376); (398-408); (477-486); 1976 (766-774).

8. Testungseffekt

ZEPP: 1973 (253-262); 1976 (117-134).

9. Trainingsprogramm

FEU: 1973 (279-288); 1975 (161-171); 1976 (328-336).

ZEPP: 1976 (274-287).

10. Unterrichtsmethode

SP: 1972 (46-54); (88-98); (227-233); (257-268); (379-384);

Anhang 2:(Fortsetzung)

FEU:1974(303-309); 1976 (283-288).

ZEPP: 1972 (1-16); (105-117); 1974 (109-123); 1975 (203-213); 1976 (117-134); (220-223).

ZeF:1975 (164-180).

DS:1973 (224-234).

11. Unterrichtsmittel

SP: 1972 (80-88); FEU: 1973 (223-234); 1974 (22-34); (35-43); 1976 (21-28); (355-363).

ZEPP: 1972 (1-16); 1973 (91-105); (279-292); 1974 (192-206); 1976 (44-50); (58-61); (235-244).

ZeF: 1974 (82-103).

DS: 1972 (398-408).

12. Unterrichtsstil

FEU: 1975 (284-294).

13. Verhaltensmodifikation

FEU: 1973 (77-88); 1974 (15-21); (249-253); 1975 (129-136); 1976 (52-59); (110-115).

ZEPP: 1972 (249-260); 1973 (16-28); (62-68).

Anhang 3: Seitenangaben der ausgewählten Artikel, in denen die abhängigen Variablen untersucht worden sind (für Untersuchung I).

1. Einstellungsänderung

PEU: 1974 (22-34); (164-175); 1975 (129-136).

DS: 1972 (398-408); 1976 (766-774).

2. Emotionale Reaktion

SP: 1972 (300-309); PEU: 1973 (77-88); 1974 (15-21); 1975 (129-136); (161-171); (284-294); 1976 (131-136).

ZEPP: 1972 (249-260); 1973 (62-68); (293-302).

3. Imitationsverhalten

ZEPP: 1972 (221-234); 1974 (75-89); 1975 (73-87).

4. Intelligenzleistung

SP: 1972 (80-88); (257-268); (268-277); PEU: 1973 (223-234); (297-307).

ZEPP: 1974 (109-123); 1976 (117-134); (274-287).

DS: 1974 (838-850).

5. Kreativität

ZEPP: 1973 (253-262); 1974 (262-280); 1976 (220-223).

DS: 1973 (224-234).

6. Lehrerverhalten

PEU: 1976 (328-336).

ZEPP: 1975 (203-213).

7. Motivationsänderung

PEU: 1973 (223-234); 1974 (22-34).

ZeF: 1975 (164-180).

8. Psychomotorik

PEU: 1974 (35-43).

ZEPP: 1973 (279-292); (293-302).

9. Regelverstöße

ZEPP: 1973 (16-28).

10. Schulleistung

SP: 1972 (46-54); (80-88); (88-98); (227-233); (379-384);

PEU: 1973 (14-30); (156-162); (223-234); (235-243);

(327-336); 1974 (15-21); (22-34); (293-302); (303-309);

1975 (284-294); 1976 (21-28); (131-136); (144-152);

(215-223); (242-245); (246-249); (283-288); (351-354);

(355-363).

Anhang 3 (Fortsetzung)

ZEPP: 1972 (1-16); (105-117); 1973 (91-105); 1974 (192-206);
1976 (37-43); (58-61); (135-141); (274-287);

ZeF: 1974 (82-103); 1975 (164-180).

DS: 1972 (362-376); (398-408); (477-486); 1974 (838-850);
1976 (766-774).

11. Soziales Verhalten

FEU: 1973 (77-88); 1974 (249-253); 1976 (52-59); (110-
115); (242-245); (351-354).

ZEPP: 1972 (151-164); 1973 (167-181); (298-308).

12. Spielverhalten

FEU: 1973 (279-288); 1976 (121-125); (173-176).

13. Wahrnehmungsfähigkeit

SP: 1972 (80-88); FEU: 1973 (77-88); 1975 (161-171).

ZEPP: 1973 (1-15); 1974 (225-240); 1976 (44-50); (235-244).

Anhang 4: Die Seitenangaben der ausgewählten Artikel aus den vier Zeitschriften für Untersuchung II.

Zeit- schriften	JABA					
	(1968)	(1970)	(1972)	(1973)	(1974)	(1976)
Jahrgänge	1-12	85-96	131-137	541-561	591-598	239-245
	35-45	97-109	139-149	563-570	599-610	247-252
	47-53	117-122	151-155	573-578	611-621	253-275
	55-61	133-138	157-162	579-586		277-288
und	73-76	139-147	163-175	621-630	(1975)	289-300
	121-129	153-157	177-182	631-642	1-12	301-314
	131-137	159-167	183-190	643-654	13-26	315-322
	139-150	169-174	225-232	671-677	53-57	323-333
Seiten- nummer	151-158	175-182	233-246		59-66	335-354
	167-173	183-184	247-262	(1974)	67-75	355-370
	213-223	199-203	263-270	1-9	77-82	377-386
	225-235	205-211	293-309	11-21	125-135	407-415
ausge- wählten	237-243	221-222	329-333	33-38	137-146	417-431
	245-250	247-255	405-420	45-59	157-168	433-444
	253-262	261-270	421-429	61-69	169-180	445-457
	267-281	273-287	467-475	71-76	187-196	459-470
Artikel	283-295	293-297	485-497	77-86	197-201	471-482
	297-306	299-307		87-91	203-209	491-497
	307-314	309-314	(1973)	103-114	279-286	
	315-322		1-14	115-121	287-299	
	329-333	(1971)	15-30	123-135	301-309	
		1-9	31-47	173-190	311-319	
	(1969)	45-59	49-56	191-198	321-330	
	3-13	77-87	57-64	199-206	353-365	
	23-29	101-112	77-78	207-215	373-380	
	31-37	121-125	79-87	233-241	381-386	
	43-47	141-149	89-100	243-256	387-398	
	49-53	163-171	101-104	257-262	399-409	
	65-77	183-189	105-113	287-297	411-420	
	85-92	191-199	115-124	299-306	421-433	
	101-112	201-213	125-130	313-325	449-457	
	113-118	215-225	193-200	357-364		
	119-124	241-247	209-217	377-383	(1976)	
	139-141	249-253	219-229	385-390	1-11	
	143-157	257-264	269-275	391-402	25-30	
	159-170	265-281	277-287	403-411	31-40	
	171-179	283-290	289-298	413-425	41-54	
	181-194	299-304	299-310	427-437	55-64	
	239-246		311-329	439-445	65-78	
	265-276	(1972)	355-366	447-460	79-84	
		7-17	397-403	461-471	85-104	
	(1970)	19-30	419-423	473-479	117-125	
	1-4	45-51	443-455	505-519	127-139	
	13-28	53-64	457-463	521-528	141-145	
	29-37	67-72	465-474	529-546	179-188	
	39-46	73-78	475-480	547-555	189-198	
	47-55	79-84	481-486	557-563	199-206	
	57-64	85-91	487-495	567-576	221-226	
	77-84	125-130	497-507	583-590	227-230	

Anhang 4. (Fortsetzung)

Zeit- schriften	<u>JBTEP</u>		<u>BT</u>		<u>BRT</u>
	(1970)	(1975)	(1970)	(1976)	(1971)
	53-58	279-282	371-381	76-88	375-379
	83-85	289-294		96-103	
Jahrgänge	257-262	295-299	(1971)	155-166	(1972)
und		339-342	78-82	240-246	411-415
Seiten-	(1971)	351-355	340-344	247-251	
nummer	145-149		375-384	348-354	(1973)
der	173-179	(1976)	519-533	382-389	179-185
ausge-	267-272	17-21	534-537	390-396	327-329
wählten	285-289	35-37		397-404	
Artikel		67-69	(1972)	472-480	(1974)
	(1972)	155-160	95-99	544-548	9-18
	31-37	161-166	349-351	602-613	107-115
	205-207	167-170	567-578		181-190
	217-222	185-187	597-603		259-260
	281-287	191-192	622-626		261-263
	297-300	227-230			327-334
		269-275	(1973)		335-345
	(1973)	355-357	22-40		
	19-30		83-90		(1975)
	161-165		123-127		45-50
	167-170		289-294		113-125
	201-207		386-391		153-162
	237-242		392-397		215-220
	243-247		431-441		
	283-287		551-558		(1976)
	389-393		570-575		53-61
					169-173
	(1974)		(1974)		391-398
	7-12		33-47		
	13-18		93-106		
	59-63		107-111		
	119-127		203-209		
	271-274		255-260		
	291-295		481-493		
			614-623		
	(1975)		(1975)		
	17-21		153-164		
	23-25		220-229		
	39-42		310-314		
	43-48		367-377		
	61-63		411-413		
	135-137		535-542		
	139-143		609-615		
	230-234		616-626		
	242-245		639-648		
	246-247				
	248-249				
	253-255		(1976)		
	264-266		54-67		

Anhang 5: Die Seitenangaben der Artikel, in denen die verschiedenen Kategorien der unabhängigen Variablen untersucht wurden (für Untersuchung II)

1. Primäre Verstärkung

JABA, (1968: 151-158; 267-281; 297-306; 307-314), (1969: 31-37) (1970: 57-64; 133-138; 293-297; 299-307; 309-314), (1971: 163-171; 183-189), (1972: 467-475), (1973: 457-463; 487-495), (1974: 257-262; 313-325; 505-519), (1975: 187-196).
JBTEP, (1970: 83-85), (1973: 389-393), (1975: 39-42), (1976: 355-357).
BT, (1972: 622-626), (1974: 93-106; 614-623), (1976: 247-251)

2. Soziale Verstärkung

JABA, (1968: 1-12; 35-45; 47-53; 55-61; 73-76; 121-129; 139-150; 167-173; 253-262; 267-281; 315-322), (1969: 3-13; 101-112; 159-170; 181-194; 239-246), (1970: 117-122; 205-211; 247-255), (1971: 121-125; 141-149; 241-247), (1972: 73-78; 79-84; 125-130; 157-162; 163-175), (1973: 15-30; 49-56; 57-64; 71-78; 115-124; 125-130; 193-200; 209-217; 475-480; 497-507; 621-630) (1974: 77-86; 257-262; 583-590), (1975: 77-82; 279-286; 449-457), (1976: 31-40; 253-275; 407-415).
JBTEP, (1970: 53-58), (1973: 19-30; 167-170), (1975: 246-247).
BT, (1972: 349-351; 567-578), (1973: 386-391; 431-441), (1974: 93-106), (1975: 153-164; 411-413), (1976: 76-88; 544-548; 602-613).

3. Tokensystem

JABA, (1968: 55-61; 151-158; 245-250), (1969: 3-13; 101-112), (1970: 47-55; 85-96; 97-109; 169-174; 199-203), (1971: 45-59; 141-149; 265-281; 299-304), (1972: 7-17; 19-30; 53-64; 73-78; 125-130; 177-182; 233-246; 247-262; 263-270; 293-309; 329-333; 421-429; 485-497), (1973: 125-130; 193-200; 219-229; 289-298; 311-329; 475-480; 541-561; 563-570; 579-586; 621-630; 671-677), (1974: 1-9; 71-76; 191-198; 207-215; 377-383; 391-402; 461-471; 567-576), (1975: 373-380; 381-386), (1976: 1-11; 79-84; 85-104; 189-198; 253-275; 277-288; 315-322; 445-457).
JBTEP, (1970: 53-58), (1971: 145-149; 173-179), (1972: 31-37; 205-207), (1973: 19-30; 167-170), (1974: 13-18), (1975:

Anhang 5 (Fortsetzung)

43-48; 61-63; 289-294; 295-299), (1976: 185-187; 227-230)
BT, (1971: 78-82; 340-344), (1972: 95-99; 567-578), (1973: 22-40; 386-391; 392-397), (1975: 153-164; 367-377; 616-626; 639-648), (1976: 76-88; 240-246).
BRT, (1973: 179-185), (1974: 181-190; 261-263), (1975: 113-125), (1976: 169-173).

4. Aktivitätskonsequenz

JABA, (1969: 43-47; 85-92; 113-118; 139-141; 171-179), (1970: 13-28; 175-182; 183-184; 247-255), (1971: 77-87; 163-171), (1972: 45-51, 151-155), (1973: 78-87; 105-113; 269-275; 419-423; 457-463), (1974: 43-38; 505-519), (1976: 417-431).

BRT, (1971: 375-379).

5. Feedback

JABA, (1968: 131-137), (1969: 85-92), (1970: 1-4), (1971: 191-199), (1972: 79-84; 139-149), (1973: 89-100; 397-403), (1974: 199-206; 413-425; 547-555; 591-598), (1975: 59-66; 449-457), (1976: 41-54; 445-457).

JBTEP, (1972: 217-222), (1974: 59-63; 271-274).

BT, (1974: 107-111), (1976: 96-103).

BRT, (1975: 45-50).

6. Selbstkontrolle

JABA, (1973: 105-113), (1974: 299-306), (1975: 387-398).

JBTEP, (1973: 243-247), (1974: 7-12), (1975: 230-234), (1976: 35-37; 161-166).

BT, (1970: 371-381), (1974: 255-260).

7. Training

JABA, (1969: 65-77), (1970: 139-147; 153-157; 273-287), (1971: 101-112; 283-290), (1974: 45-59; 123-135; 357-364; 427-437; 439-445; 447-460; 473-479; 599-610; 611-621), (1975: 67-75; 125-135; 157-168; 169-180; 301-309; 311-319; 321-330; 399-409; 421-433), (1976: 65-78; 85-104; 117-125; 127-139; 179-188; 239-245; 323-333; 335-354; 355-370; 433-444; 459-470).

JBTEP, (1972: 297-300), (1974: 291-295), (1975: 17-21; 339-342).

BT, (1971: 375-384; 519-533), (1973: 288-294; 551-558), (1976: 348-354; 382-389; 397-404).

BRT, (1974: 9-18), (1976: 391-398).

Inhalt - (Fortsetzung)

6. Verstärkungszug

IASA, (1969: 119-124; 265-276), (1971: 45-59), (1972: 53-64; 125-130; 293-309; 329-333; 485-497), (1973: 219-229), (1974: 173-190; 403-411; 567-576), (1975: 287-299), (1976: 2-10).

BSSE, (1976: 191-192; 227-230).

BT, (1971: 534-537), (1976: 76-88).

9. Ignorieren

IASA, (1968: 139-150), (1969: 3-13), (1972: 53-64).

BSSE, (1973: 3-10).

BT, (1973: 56-59), (1975: 570-575).

10. Darbietung aversiver Reize

IASA, (1969: 143-157), (1971: 201-213), (1974: 173-190; 557-563), (1975: 13-26; 53-57).

BSSE, (1973: 383-387), (1975: 242-245; 246-247; 248-249; 264-266; 351-355), (1976: 155-160).

BT, (1974: 614-623).

BSZ, (1972: 411-415), (1973: 327-329).

11. Ausweichverhalten

IASA, (1968: 47-53), (1969: 31-37; 159-170; 181-194; 239-246), (1970: 17-21), (1971: 255-275; 341-347), (1972: 67-72; 85-91; 163-175), (1973: 79-87; 101-104; 289-298; 443-455), (1974: 173-190; 313-325), (1976: 55-64; 471-482).

BSSE, (1973: 33-38), (1975: 231-242), (1976: 139-143; 248-249; 253-255; 295-299), (1976: 17-21; 269-275; 355-357).

BT, (1973: 123-127; 431-441), (1974: 481-493), (1975: 609-615), (1976: 247-251; 544-548).

BSZ, (1974: 107-115; 327-334), (1976: 53-61).

12. Diskontinuität

IASA, (1973: 1-14), (1974: 385-390).

BT, (1976: 390-396).

BSZ, (1975: 153-162), (1976: 53-61).

13. Unterbrechung

IASA, (1974: 521-528), (1975: 1-11).

BSSE, (1976: 167-170).

BSZ, (1976: 53-61).

14. Maßgabe

IASA, (1968: 139-150), (1969: 3-13), (1971: 45-59).

15. Transfer

Anhang 5 (Fortsetzung)

JABA, (1969:3-13), (1970:159-167), (1971:113-119), (1972:467-475), (1973: 89-100; 419-423), (1976: 55-64; 407-415).

JBTEP, (1975:135-137), (1976:67-69).

BT, (1976:472-480).

BRT, (1974: 181-190; 335-345).

16.Prompting

JABA, (1969:23-29), (1970:221-222), (1972:131-137), (1974:115-121), (1975:197-201), (1976:31-40; 141-145; 315-322; 377-386; 471-482).

17.Imitation

JABA, (1973:299-310), (1974:233-241), (1975:203-209).

JBTEP, (1975:135-137).

BT, (1976:382-389).

18.Trainingsgeräte

JABA, (1968:283-295), (1971:249-252).

BRT, (1975:215-220).

19.Systematische Desensibilisierung

JBTEP, (1970:257-262), (1971: 267-272; 285-289).

BT, (1972:597-603).

20.Beschäftigungstherapie

BT, (1976:247-251).

21.Orgastische Rekonditionierung

BT, (1976:155-166).

22.Entspannung

JBTEP, (1976:67-69).

BT, (1974:203-209).

23.Gesprächstherapie

JBTEP, (1976:155-160).

24.Stimuluskontrolle

JBTEP, (1975:279-282).

25.Wecken

JBTEP, (1975:23-25).

26.Negative Übung

JBTEP, (1972:281-287)

BT, (1975:535-542).

BRT, (1974:259-260).

27.Sonstige

JABA, (1968: 213-223; 225-235; 237-243; 329-333), (1969:23-29;

Anhang 5.(Fortsetzung)

49-53; 65-77), (1970: 29-37; 39-46; 261-270), (1971: 1-9; 257-264), (1972: 183-190; 225-232; 405-420), (1973: 31-47; 277-287; 355-366; 465-474; 481-486; 573-578; 621-630; 631-642; 643-654), (1974: 11-21; 61-69; 87-91; 103-114; 243-256; 287-297; 413-425; 529-546), (1975: 137-146; 353-365; 411-420), (1976: 199-206; 221³-226; 227-230; 247-252; 289-300; 301-314; 323-333; 491-497).

JBTEP, (1973:201-207), (1974:119-127).

BT, (1973:83-90), (1974:33-47), (1975:220-229; 310-314), (1976:54-67).

Anhang 6: Die Seitenangaben der Artikel, in denen die verschiedenen Kategorien der abhängigen Variablen untersucht wurden (für Untersuchung II)

1. Aufmerksamkeitsverhalten

JABA, (1968: 1-12; 55-61; 245-250), (1969:85-92), (1970: 13-28; 97-109; 205-211; 247-255; 293-297), (1971: 191-199; 257-264), (1972: 1-17; 79-84; 151-155), (1973: 15-30; 71-78; 105-113; 419-423; 465-474; 643-654), (1974: 115-121; 233-241; 299-306; 313-325; 413-425), (1975: 187-196; 203-209; 279-286; 373-380; 449-457), (1976: 179-188).

JBTEP, (1973:237-242).

BT, (1970:371-381), (1972: 349-351; 622-626), (1975: 153-164; 310-314), (1976:76-88).

2. Pünktlichkeit

JABA, (1970:247-255), (1971:45-59), (1973:563-570), (1976: 221-226).

3. Sauberkeit und Ordnung

JABA, (1968:213-223), (1971:45-59), (1973:541-561; 579-586), (1974:377-383), (1975: 301-309; 381-386), (1976:253-275; 277-288; 315-322; 417-431).

JBTEP, (1972:31-37), (1974:13-18), (1975:289-294), (1976:227-230)

BT, (1971:340-344), (1973:22-40).

BRT, (1974:181-190), (1975:113-125).

4. Prosoziale Verhaltensweisen

JABA, (1968: 73-76; 121-129; 167-173; 307-314), (1969:239-246), (1970: 133-138; 169-174; 309-314), (1971:241-247), (1972: 125-150; 163-175), (1973: 193-200; 573-578), (1974: 583-590; 599-610), (1976: 31-40; 65-78; 335-354; 491-497).

JBTEP, (1970:53-58), (1975:39-42), (1976:35-37).

BT, (1972:567-578), (1973:386-391), (1975:220-229), (1976: 348-354; 602-613).

5. Partizipation

JABA, (1970: 29-37; 57-64; 175-182), (1972: 225-232; 405-420), (1973:31-47), (1974: 77-86; 199-206; 391-402), (1976: 141-145; 199-206).

JBTEP, (1973:201-207), (1974:119-127), (1975:43-48).

BT, (1971:534-537), (1975:639-648), (1976:390-396; 602-613).

Anhang 6 (Fortsetzung)

BRT, (1975:113-125).

6. Aufgabenausführen

JABA, (1969: 43-47; 49-53; 139-141), (1970:47-55), (1971:299-304), (1972: 53-64; 263-270), (1974: 199-206; 207-215; 547-555), (1975:59-66), (1976: 41-54; 85-104; 277-288; 417-431).

JBTEP, (1972:297-300).

7. Soziale Fertigkeit

JABA, (1975:125-135), (1976:127-139; 239-245; 301-314; 377-386).

JBTEP, (1974:291-295), (1975:135-137).

BT, (1973:551-558), (1976:582-589).

BRT, (1976:391-398).

8. Techniken zur Verhaltensmodifikation

JABA, (1968: 47-53; 315-322), (1970: 1-4; 153-157), (1972: 19-30; 139-149), (1973: 89-100; 497-507), (1974: 11-21; 103-114; 427-437), (1975: 169-180; 197-201; 311-319; 373-380), (1976: 355-370; 459-470).

JBTEP, (1975:539-542).

BT, (1971:519-533).

9. Störverhalten

JABA, (1968: 35-45; 139-150; 245-250), (1969: 3-13; 31-37; 113-118; 119-124; 171-179), (1970: 47-55; 183-184; 293-297), (1971: 113-119; 141-149; 191-199), (1972: 45-51; 73-78; 293-309; 329-333; 485-497), (1973: 15-30; 49-56; 79-87; 269-275; 277-287; 457-463; 671-677), (1974: 71-76; 287-297; 403-411; 567-576; 591-598), (1975: 137-146; 287-299; 421-433), (1976: 55-64; 199-206; 445-457).

JBTEP, (1971:173-179), (1973:237-242), (1975: 253-255; 295-299), (1976: 17-21; 185-187; 355-357).

BT, (1973: 83-90; 431-441), (1974:107-111), (1975: 153-164; 310-314; 609-615; 616-626), (1976:240-246; 602-613).

10. Ungehorsamkeit

JABA, (1969: 159-170; 181-194), (1973: 125-130; 289-298), (1974:357-364), (1975:77-82), (1976:471-482).

JBTEP, (1973:19-30).

BT, (1975:609-615), (1976: 544-548; 602-613).

11. Aggressives Verhalten

JABA, (1968: 213-223; 237-243), (1969:31-37), (1970:199-203), (1971:215-225), (1972: 53-64; 85-91), (1973: 115-124;

Anhang 6 (Fortsetzung)

443-455), (1974:313-325), (1976: 55-64; 117-125; 189-198).

JBTEP, (1972:205-207), (1973: 19-30; 161-165), (1975: 139-143; 295-299).

BT, (1974:481-493), (1975:609-615), (1976:602-613).

12.Sprachfähigkeit

JABA, (1968: 151-158; 237-243; 253-262; 267-281; 283-295; 297-306), (1969:265-276), (1970: 39-46; 85-96; 273-287; 299-307), (1971:265-281), (1972:183-190; 233-246), (1973: 219-229; 249-310; 311-329; 397-403; 621-630), (1974: 243-256; 439-445; 447-460; 461-471; 473-479; 611-621), (1975: 321-330; 411-420), (1976:355-370).

BT, (1974:93-106), (1976: 96-103; 397-404).

13.Schulaufgabe

JABA, (1968:329-333), (1970: 139-147; 183-184; 247-255; 261-270), (1971: 77-87; 163-171), (1972: 7-17; 79-84; 157-162; 177-182), (1973: 475-480; 481-486; 631-642), (1974: 61-69; 71-76; 77-86; 87-91; 115-121; 505-519; 567-576), (1975:67-75; 157-168; 387-398; 449-457), (1976: 199-206; 227-230; 247-252; 289-300; 323-333).

JBTEP, (1972:217-222), (1976:191-192).

BT, (1972:567-578), (1975: 367-377; 639-648).

BRT, (1976:169-173).

14.Kreatives Verhalten

JABA, (1972:421-429), (1973:209-217), (1976:79-84).

15.Vorschulisches Verhalten

JABA, (1968:225-235), (1969:101-112), (1970: 117-122; 159-167; 221-222), (1971: 1-9; 101-112; 283-290), (1972:467-475), (1973: 79-87; 497-507), (1974:45-59; 123-135).

JBTEP, (1973:389-393; 1975:17-21).

16.Ausscheidungsverhalten

JABA, (1971: 241-247; 249-253).

JBTEP, (1970:83-85), (1975:23-25).

BT, (1973:288-294).

BRT, (1975:215-220).

17.EBverhalten

JABA, (1969:65-77), (1970:77-84), (1972:67-72), (1976:407-415).

JBTEP, (1975:246-247).

BT, (1975:535-542).

18.Motorik

Anhang 6 (Fortsetzung)

JABA, (1971:183-187), (1972:131-137), (1975:399-409).

BT, (1971:375-384), (1975:411-413).

19.Selbstdestruktives Verhalten

JABA, (1969:143-157), (1971:201-213), (1975:53-57).

JBTEP, (1974:7-12).

BT, (1974:614-623).

20.Sexualstörungen

JABA, (1973:355-366), (1974:173-190).

JBTEP, (1970:257-262), (1971:285-289), (1974:271-274).

BT, (1974:33-47), (1976:155-166).

BRT, (1972:411-415), (1974:335-345), (1975:45-50), (1976:53-61).

21.Ubermäßige Phantasie und bizarre Sprechweise

JABA, (1972:247-262), (1973:57-64).

JBTEP, (1971:145-149), (1973:167-170), (1974:13-18), (1975:
61-63), (1976:155-160; 269-275).

BT, (1973: 123-127; 570-575).

22.Phobien

JABA, (1968:131-137).

BT, (1972:597-603).

23.Fallsucht

JABA, (1975:1-12).

BT, (1976:247-251).

24.Selbst-Stimulation

JABA, (1971:121-125), (1973:1-14; 487-495), (1974: 33-38;
385-390; 521-528).

JBTEP, (1973:237-242), (1975:248-249).

25.Sucht

JABA, (1975:13-26).

JBTEP, (1973:383-387), (1974:7-12), (1975:242-245).

BT, (1976:472-480).

BRT, (1971:375-379), (1974:107-115; 261-263; 327-334).

26.Amnesie

JBTEP, (1975:351-355), (1976:67-69).

27.Depressives Verhalten

JABA, (1972:125-150).

JBTEP, (1973:243-247), (1975:279-282).

BT, (1973:392-397).

28.Scavenging behavior

BRT, (1975:153-162).

Anhang 6 (Fortsetzung)

29. Zucker.

JBTEP, (1972:281-287), (1976:167-170).

RT, (1974:259-260).

30. Zwangsverhalten

JBTEP, (1971:267-272), (1975:264-266).

BT, (1974:255-260).

31. Hypertension

JBTEP, (1974:59-63), (1975:230-234).

BT, (1974:203-209), (1976:54-67).

BRT, (1973:179-185).

32. Sonstige

JABA, (1969:23-29), (1973:101-104), (1974: 1-9; 191-198; 257-262; 529-546; 557-563), (1975:353-365), (1976: 1-11; 23-30; 433-444).

JBTEP, (1976:161-166).

BT, (1971:78-82), (1972:95-99).

BRT, (1973:327-329), (1974:9-18).

Anhang 7: Seitenangaben der Artikel, in denen die verschiedenen
Einzelfall-Experimentaldesigns angewendet wurden

11: ABAB-Design

JABA, (1968: 1-12; 47-53; 55-61; 73-76; 121-129; 139-150;
151-158; 167-173; 213-223; 225-235; 237-243; 245-250;
253-262; 283-295; 297-306; 307-314; 315-322; 329-333),
(1969: 3-13; 31-37; 43-47; 49-53; 65-77; 113-118; 139-
141; 159-170; 171-179; 181-194; 265-276), (1970: 13-28;
29-37; 47-55; 57-64; 97-109; 117-122; 133-138; 159-167;
175-182; 183-184; 199-203; 221-222; 261-270; 293-297;
299-307; 309-314), (1971: 1-9; 113-119; 121-125; 141-149;
183-189; 191-199; 241-247; 249-253; 257-264; 283-290),
(1972: 7-17; 45-51; 53-64; 67-72; 73-78; 85-91; 125-130;
131-137; 139-149; 157-162; 183-190; 225-232; 247-262;
263-270; 293-309; 329-333; 405-420), (1973: 1-14; 15-30;
31-47; 49-56; 71-78; 101-104; 209-217; 219-229; 277-287;
299-310; 355-366; 397-403; 457-463; 475-480; 541-561;
563-570; 573-578; 579-586; 643-654), (1974: 33-38; 61-
69; 71-76; 87-91; 115-121; 173-190; 191-198; 199-206;
207-215; 233-241; 257-262; 287-297; 299-306; 313-325;
377-383; 461-471; 473-479; 505-519; 521-528; 529-546;
547-555; 557-563; 567-576; 583-590; 591-598), (1975:
1-12; 53-57; 287-299; 373-380), (1976: 25-30; 31-40;
55-64; 141-145; 199-206; 221-226; 227-230; 247-252;
253-275; 277-288; 301-314; 315-322; 323-333; 445-457;
491-497).

JBTEP, (1970: 53-58; 83-85; 257-262), (1971: 145-149; 173-179;
285-289), (1972: 31-37; 205-207; 217-222; 297-300),
(1973: 19-30; 161-165; 167-170; 201-207; 237-242; 383-
387; 389-393), (1974: 59-63; 119-127; 271-274), (1975:
39-42; 43-48; 61-63; 242-245; 248-249; 264-266; 279-282;
295-299; 351-355), (1976: 35-37; 67-69; 161-166; 167-170;
185-187; 227-230).

BT, (1970: 371-381), (1971: 78-82; 340-344; 375-384; 519-533;
534-537), (1972: 95-99; 349-351; 622-626), (1973: 22-40;
123-127; 288-294; 392-397; 551-558), (1974: 33-47; 93-
106; 107-111; 203-209), (1975: 367-377; 535-542; 609-615;
616-626; 639-648), (1976: 54-67; 96-103).

BRT, (1971: 375-379), (1972: 411-415), (1973: 179-185; 327-329), (1974: 107-115; 261-263; 335-345), (1975: 45-50; 113-125; 215-220), (1976: 53-61; 169-173).

12: ABACA-Design

JABA, (1973: 79-87; 105-113; 621-630; 631-642), (1974: 103-114; 529-546), (1975: 13-26; 77-82), (1976: 221-226; 471-482).

JBTEP, (1975: 139-143), (1976: 155-160).

BT, (1973: 83-90), (1976: 155-166).

BRT, (1974: 181-190; 327-334).

13: ABABB'B"-Design

JABA, (1968: 35-45), (1970: 39-46; 139-147), (1971: 45-59; 77-87), (1976: 85-104).

JBTEP, (1974: 13-18).

BT, (1975: 310-314), (1976: 247-251).

14: B-BC-B-Design

JABA, (1968: 131-137), (1969: 101-112), (1970: 29-37; 97-109) (1971: 163-171; 215-225), (1972: 7-17; 79-84; 467-475), (1973: 419-423; 481-486), (1975: 449-457).

JBTEP, (1975: 230-234), (1976: 269-275; 355-357).

BT, (1973: 386-391; 431-441), (1975: 411-413), (1976: 76-88).

BRT, (1974: 259-260).

15: {^{AB3}_{AB3}-Design

JABA, (1968: 267-281), (1969: 85-92; 119-124; 143-157; 239-246), (1970: 1-4; 77-84; 85-96; 153-157; 205-211; 247-255; 273-287), (1971: 101-112; 191-199; 201-213; 265-281; 299-304), (1972: 19-30; 163-175; 233-246; 405-420; 421-429; 485-497), (1973: 57-64; 89-100; 115-124; 125-130; 269-275; 289-298; 311-329; 443-455; 465-474; 497-507; 671-677), (1974: 11-21; 45-59; 61-69; 77-86; 115-121; 123-135; 173-190; 199-206; 243-256; 357-364; 377-383; 385-390; 391-402; 403-411; 413-425; 427-437; 439-445; 447-460; 473-479; 547-555; 599-610; 611-621), (1975: 59-66; 67-75; 125-135; 137-146; 157-168; 169-180; 187-196; 197-201; 203-209; 279-286; 301-309; 311-319; 321-330; 353-365; 381-386; 387-398; 399-409; 411-420; 421-433; 449-457), (1976: 1-11; 31-40; 41-54; 65-78; 79-84; 85-104; 117-125; 127-139; 179-188; 189-198; 239-245;

$222-222$, $222-222$, $222-222$, $222-222$, $222-222$, $222-222$,
 $222-222$, $222-222$, $222-222$.

22222, (22222) , $222-222$, $222-222$, $222-222$, $222-222$, $222-222$,
 (22222) , $222-222$, $222-222$, $222-222$, $222-222$, $222-222$,
 $222-222$, $222-222$, $222-222$, $222-222$, $222-222$, $222-222$.

222, (22222) , $222-222$, (22222) , (22222) , $222-222$, $222-222$,
 $222-222$, $222-222$, $222-222$, $222-222$, $222-222$, $222-222$,
 $222-222$, $222-222$, $222-222$, $222-222$, $222-222$, $222-222$,
 $222-222$.

2222, (22222) , $222-222$, $222-222$, $222-222$, $222-222$.

22, $\left\{ \begin{matrix} 22222 \\ 22222 \end{matrix} \right\} = 22222222$

22222, (22222) , $222-222$, (22222) , $222-222$, $222-222$, $222-222$,
 $222-222$.

22222222, (22222222) .

22, $222-222$.

Herrn Prof. Dr. V. Krumm möchte ich an dieser Stelle besonders herzlich für seine Anleitung und für seine nie ermüdende Hilfsbereitschaft beim Entstehen dieser Arbeit danken. Mein besonderer Dank gilt darüber hinaus Herrn N. Posse und Herrn P. Dürschmidt, die diese Arbeit gelesen und freundschaftliche Ratschläge gegeben haben.

Lebenslauf

Am 21. September 1944 wurde ich als Sohn des Bauern In-ho Ma und seiner Ehefrau Shue-nuan geb. Lin in Tainan, Taiwan geboren. Von 1951-1957 ging ich in die dortige Shen-li-Volksschule. Ab 1957 besuchte ich die fünfjährige Hsin-feng-Landwirtschaftsschule, wo ich im Juli 1962 die Reifeprüfung ablegte. Danach wurde ich durch eine Aufnahmeprüfung in die Nationale Chung-hsing-Universität für das Fach "landwirtschaftliche Erziehung" aufgenommen. Nach dem vierjährigen Studium und einjährigen Lehramtspraktikum erwarb ich "the degree of Bachelor of Science". Anschließend wurde ich zum einjährigen Militärdienst eingezogen. Danach immatrikulierte ich mich an der Nationalen Chengchi-Universität für das Fach "Erziehungswissenschaft". 1970 legte ich in Teipeh die Prüfung für "the degree of Master of Education". Im Sommer 1970 ging ich als Forschungsassistent zum psychologischen Institut der Nationalen Taiwan-Universität. Von Oktober 1971 bis September 1975 arbeitete ich in dem Ausbildungszentrum in der Chinesischen Petroleum Corporation. In WS 1975/76 besuchte ich einen Deutschkurs in der Universität zu Köln, wo ich die Sprachprüfung ablegte. Seit SS 1976 bin ich unter der Anleitung von Prof. Dr. V. Krumm als Doktorand tätig.

